

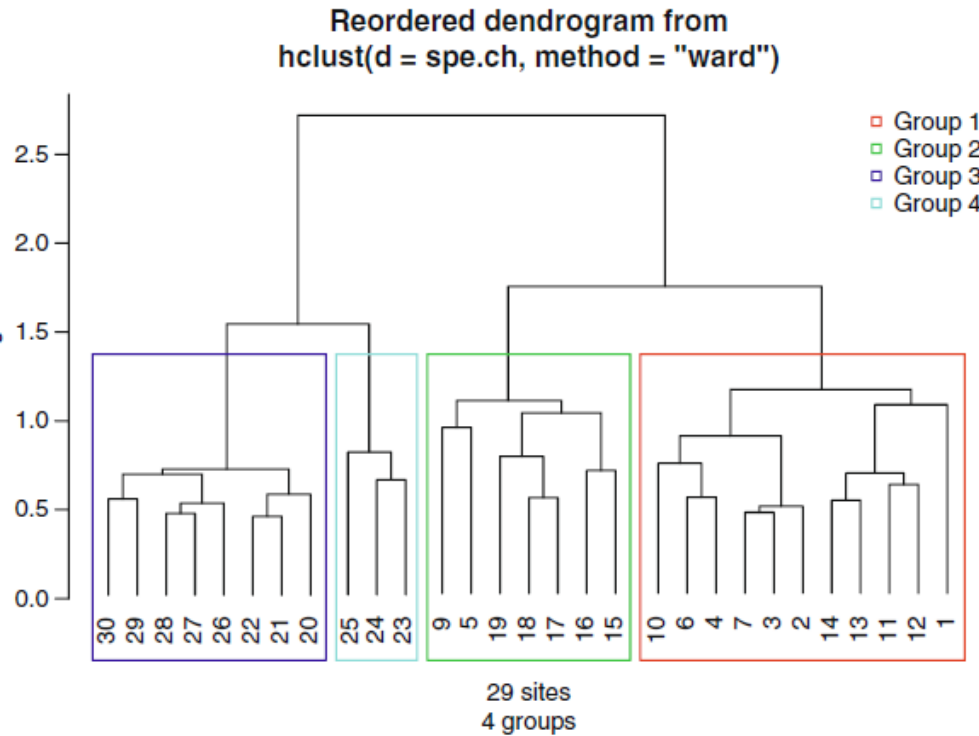
第4讲 多元统计分析

多元数据

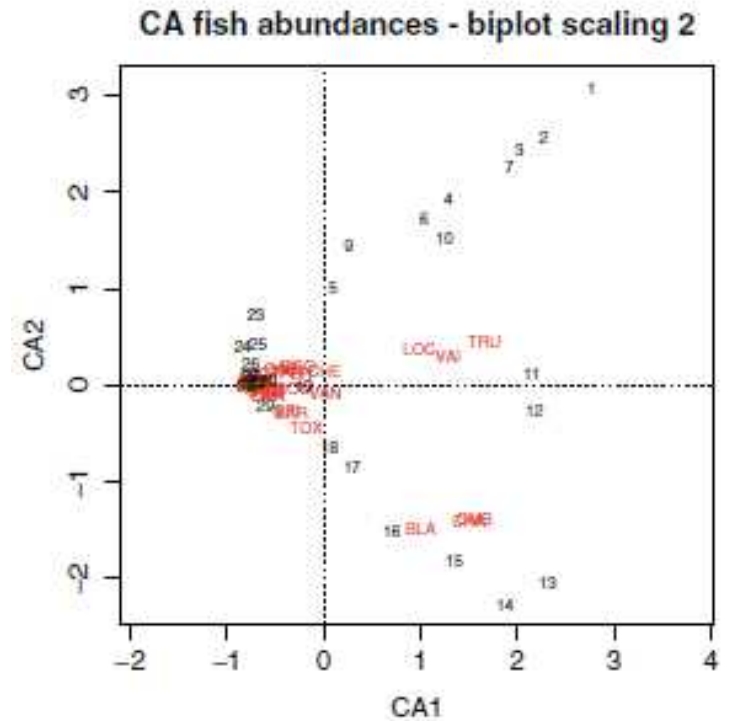
$$X = \{x_{ij}\} = \left\{ \begin{array}{ccccc} x_{11} & x_{12} & x_{13} & \square & x_{1N} \\ x_{21} & x_{22} & x_{23} & \square & x_{2N} \\ \square & \square & \square & \square & \square \\ x_{p1} & x_{p2} & x_{p3} & \square & x_{pN} \end{array} \right\}$$

$i = 1, 2, \square \quad , \quad P; \quad j = 1, 2, \square \quad N$

多元数据基本方法



聚类 (cluster)



排序 (ordination)

► 数据分析与模拟丛书

Daniel Borcard François Gillet Pierre Legendre 著
赖江山 译

Numerical Ecology with R

数量生态学

——R 语言的应用



Daniel Borcard 博士

加拿大蒙特利尔大学生物科学系高级研究员。长期从事数量生态学、群落生态学和生物统计学方面的科研和教学工作。曾获蒙特利尔大学优秀教学奖。



François Gillet 博士

法国弗朗什孔泰大学 (Université de Franche-Comté) 科学与技术学院教授。长期从事群落生态学、植被生态学、数量生态学和生态模型方面的科研和教学工作。



Pierre Legendre 博士

加拿大蒙特利尔大学生物科学系教授，加拿大皇家学会会员。在群落生态学、数量生态学、统计生态学和系统发育中的数量方法等领域具有极高的造诣，是国际数量生态学界的权威，著有影响力极大的 *Numerical Ecology* 一书，发表论文 200 余篇。

高等教育出版社

Use R!

Daniel Borcard
François Gillet
Pierre Legendre

Numerical
Ecology with R

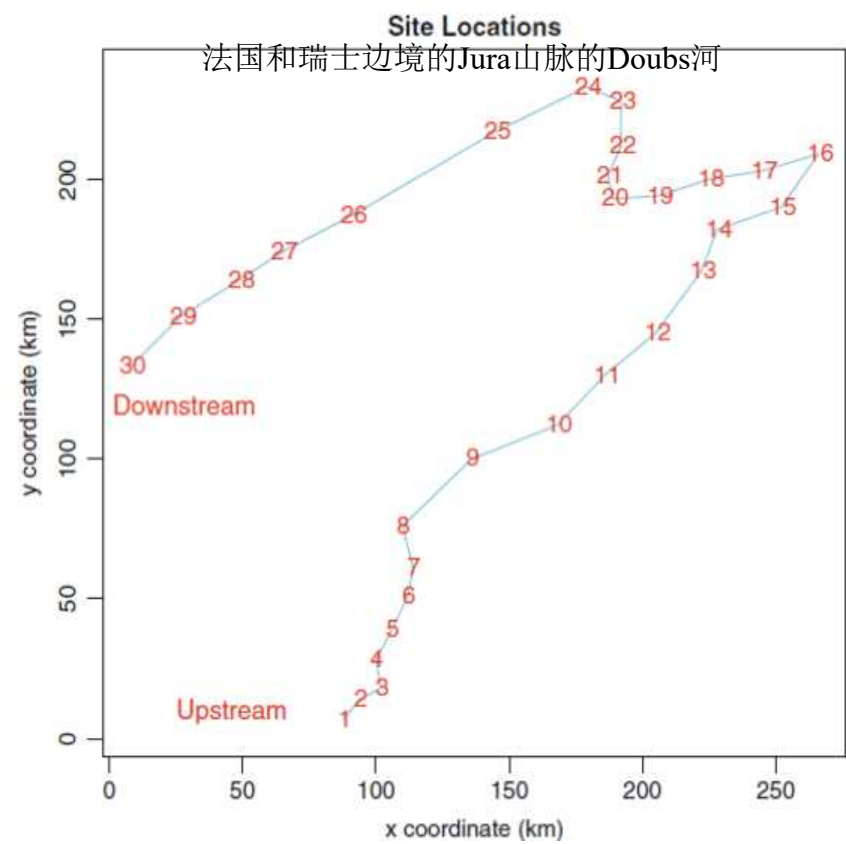
Springer



NUMERICAL ECOLOGY

Pierre Legendre &
Louis Legendre

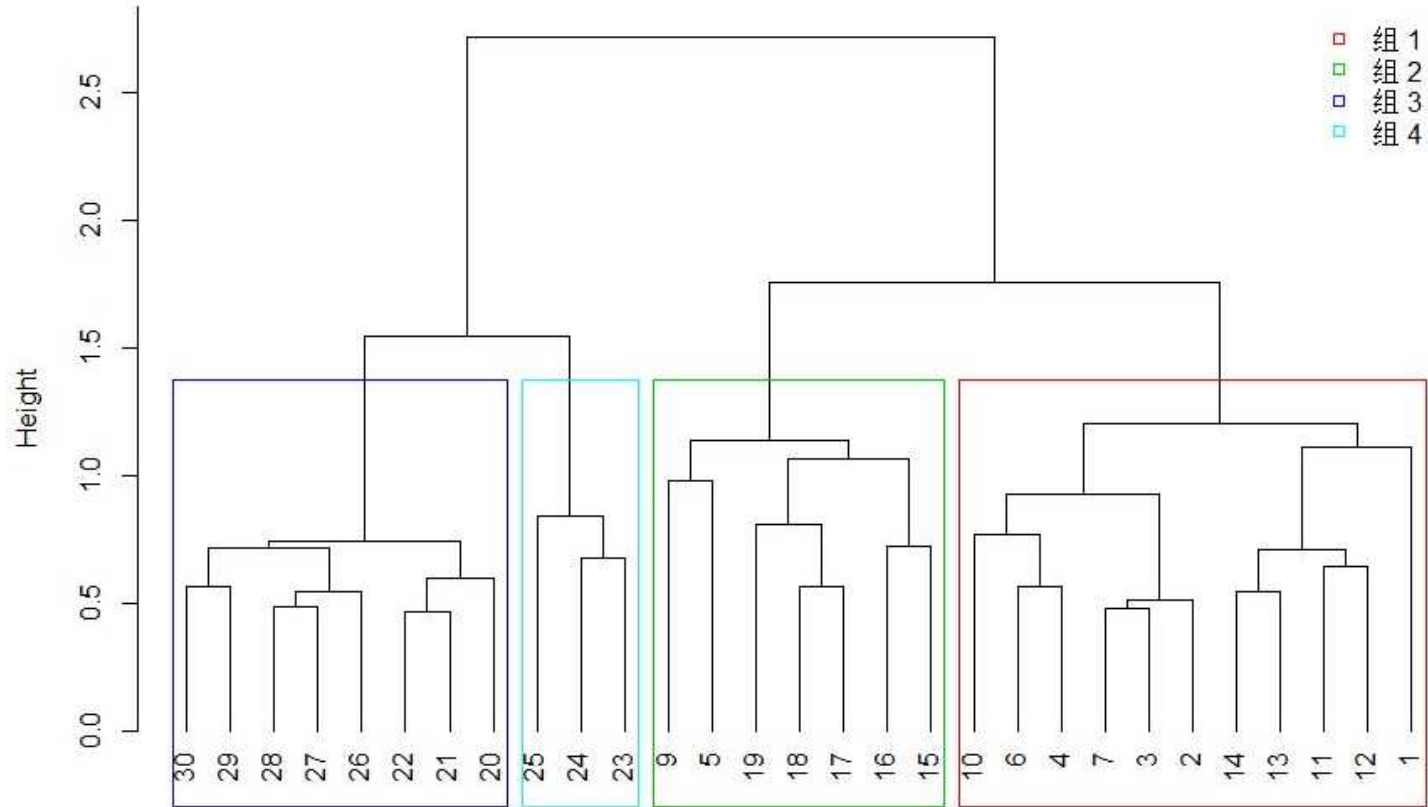
Doubs鱼类数据集



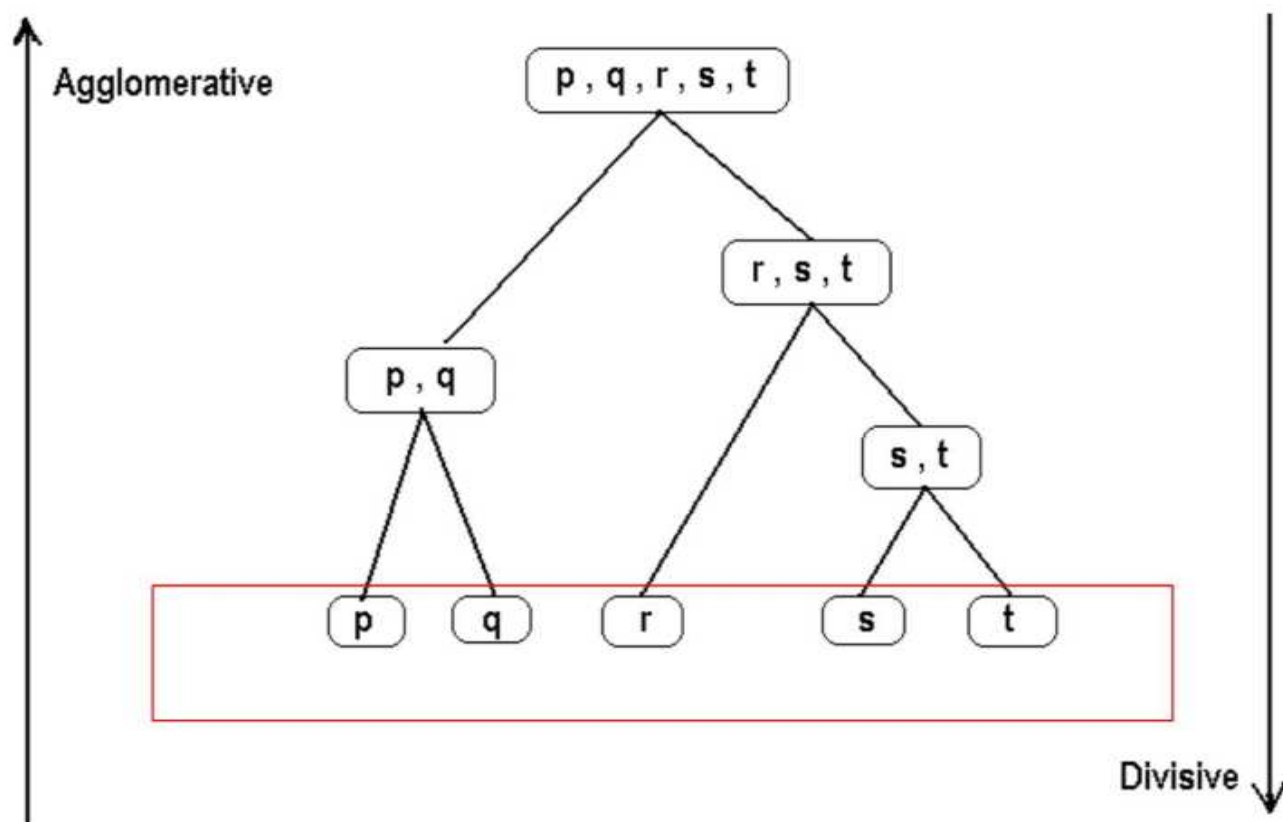
变量名称	代号	单位
离源口距离	das	km
海拔	Alt	m a.s.l.
坡度	pen	‰
平均最小流量	deb	m ³ s ⁻¹
pH值	pH	-
硬度 (毫克/升)	dur	mgL ⁻¹
磷酸盐度 (毫克/升)	pho	mgL ⁻¹
硝酸盐度 (毫克/升)	nit	mgL ⁻¹
氨氮度 (毫克/升)	amm	mgL ⁻¹
氧含量	oxy	mgL ⁻¹
生物需氧量	dbo	mgL ⁻¹

该数据集包括3个矩阵，第一个矩阵是27种鱼类在每个样方的多度，第二个矩阵包括11个与河流的水文、地形和水体化学属性相关的环境变量，第三个矩阵是样方的地理坐标（笛卡尔坐标系，X和Y）。

聚类分析



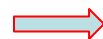
29 样方
4 组



层次法（hierarchical）和非层次法（non-hierarchical）

排序的对象-距离矩阵

Objects	Descriptors						
	y_1	y_2	y_3	...	y_j	...	y_p
x_1	y_{11}	y_{12}	y_{13}	...	y_{1j}	...	y_{1p}
x_2	y_{21}	y_{22}	y_{23}	...	y_{2j}	...	y_{2p}
x_3	y_{31}	y_{32}	y_{33}	...	y_{3j}	...	y_{3p}
...	...	...	...	...	...	...	...
...	...	...	...	...	...	...	...
...	...	...	...	...	...	...	...
x_i	y_{i1}	y_{i2}	y_{i3}	...	y_{ij}	...	y_{ip}
...	...	...	...	...	...	...	...
...	...	...	...	...	...	...	...
...	...	...	...	...	...	...	...
x_n	y_{n1}	y_{n2}	y_{n3}	...	y_{nj}	...	y_{np}



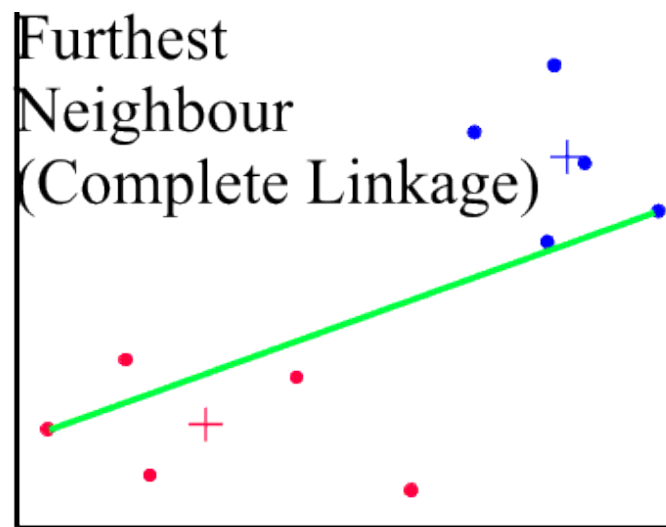
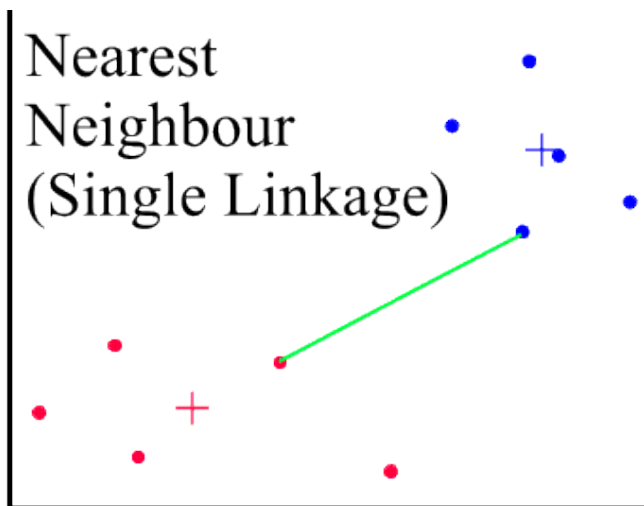
$$A_{nn} =$$

$$\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{bmatrix}$$

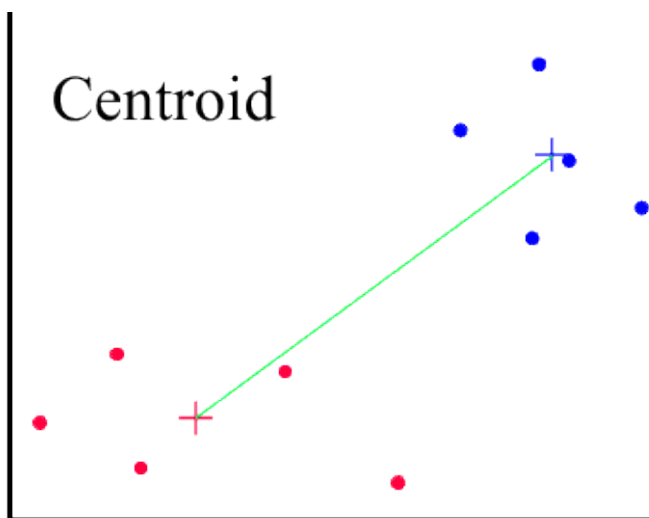
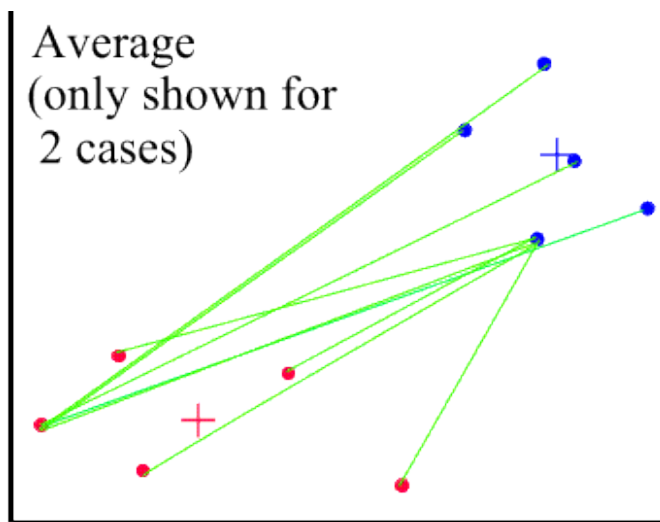
$$dist(a_1, a_2) = \sqrt{\sum_{i=1}^n (y_{1i} - y_{2i})^2}$$

距离（distance）（相异性）矩阵。
例如欧氏距离、Jaccard相似系数

基于连接的层次聚类



平均聚合聚类



#加载所需的程序包

```
library(ade4)
```

```
library(vegan) #应该先加载ade4后加载vegan以避免冲突
```

```
library(gclus)
```

```
library(cluster)
```

```
library(RColorBrewer)
```

```
library(labdsv)
```

```
library(mvpart)
```

```
library(MVPARTwrap) # MVPARTwrap这个程序包必须从本地zip文件安装
```

导入CSV格式的数据

```
spe <- read.csv("DoubsSpe.csv", row.names=1)
```

```
env <- read.csv("DoubsEnv.csv", row.names=1)
```

```
spa <- read.csv("DoubsSpa.csv", row.names=1)
```

删除无物种数据的样方8

```
spe <- spe[-8,]
```

```
env <- env[-8,]
```

```
spa <- spa[-8,]
```

#物种多度数据：先计算样方之间的弦距离矩阵，然后进行单连

#接聚合聚类

```
spe.norm <- decostand(spe, "normalize")
```

```
spe.ch <- vegdist(spe.norm, "euc")
```

```
spe.ch.single <- hclust(spe.ch, method="single")
```

```
par(mfrow=c(2,2))
```

使用默认参数选项绘制聚类树

```
plot(spe.ch.single)
```

#计算完全连接聚合聚类

```
# *****
```

```
spe.ch.complete <- hclust(spe.ch, method="complete")
```

```
plot(spe.ch.complete)
```

```
# 计算UPGMA聚合聚
```

```
# *****
```

```
spe.ch.UPGMA <- hclust(spe.ch, method="average")
```

```
plot(spe.ch.UPGMA)
```

个UPGMA聚合聚 看起来介于 接聚 和完全 接聚 之 。
种# 情况 常 生。

```
# 计算鱼类数据的形心聚类
```

```
# *****
```

```
spe.ch.centroid <- hclust(spe.ch, method="centroid")
```

```
plot(spe.ch.centroid)
```

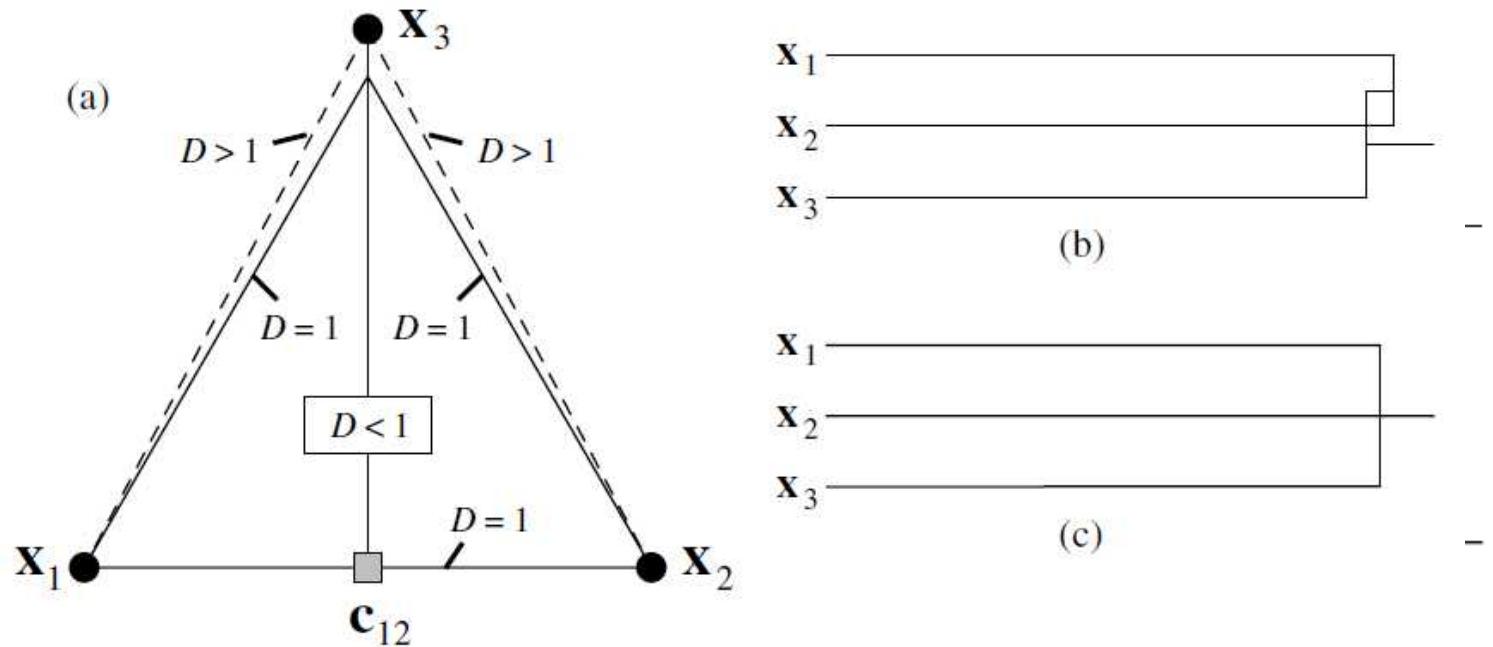
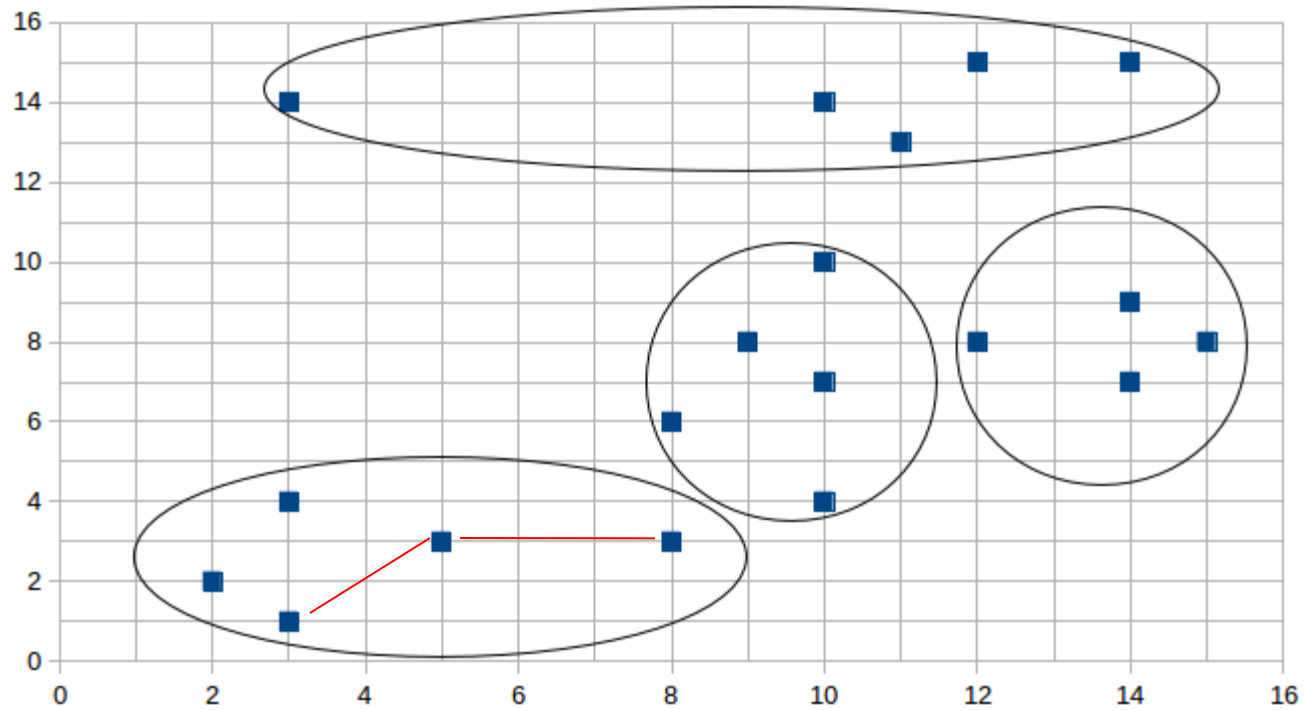


Figure 8.16 A reversal may occur in situations such as (a), where x_1 and x_2 cluster first because they represent the closest pair, although the distance from x_3 to the centroid c_{12} is smaller than the distance from x_1 to x_2 . (b) The result is usually depicted by a non-ultrametric dendrogram. (c) The reversal may also be interpreted as a trichotomy.

Ward最小方差聚类



```
# 计算Ward最小方差聚
```

```
# *****
```

```
par(mfrow=c(2,1))
```

```
spe.ch.ward <- hclust(spe.ch, method="ward")
```

```
plot(spe.ch.ward)
```

```
#使用距离平方造成此聚图上半部分过于膨胀。为了使聚图比例看起来
```

```
#更图而不影响图构，可以使用当前融合水平的平方根重新图
```

```
spe.ch.ward$height <- sqrt(spe.ch.ward$height)
```

```
plot(spe.ch.ward)
```

非层次聚类

- 非层次聚类（non-hierarchical clustering）是对一组对象进行简单分组的方法，也可以表述为：在 p 维空间内有 n 个对象（点），将 n 个对象分为 k 组（或称为聚类簇），分组的依据是尽量使组内的对象之间比组间对象之间的相似度更高。此时用户需要自己决定分组的数量 k 。非层次聚类的算法首先需要有个初始的结构，即首先将所有对象任意分为 k 组，然后在初始结构的基础上进行不断替换迭代，以达到最优化的分组结果。初始结构的设定可以依据某种理论，但大多数情况下是随机分配。通常是设定不同的初始结构，然后通过大量的迭代以找到最佳的解决方案。

#预转化后物种数据k-均值划分

*****

```
spe.kmeans <- kmeans(spe.norm, centers=4, nstart=100)
```

```
spe.kmeans
```

#注意：即使给定的**nstart**相同，每次运行上述命令，所产生的结果也不一定

#完全相同，因为每次运算设定的初始结构是随机的。

如果不知道多多少组，怎么办，下面SSI 图帮助你

k-均值划分，2组到10组

```
spe.KM.cascade <- cascadeKM(spe.norm, inf.gr=2, sup.gr=10, iter=100,  
  criterion="ssi")
```

```
plot(spe.KM.cascade, sortg=TRUE)
```

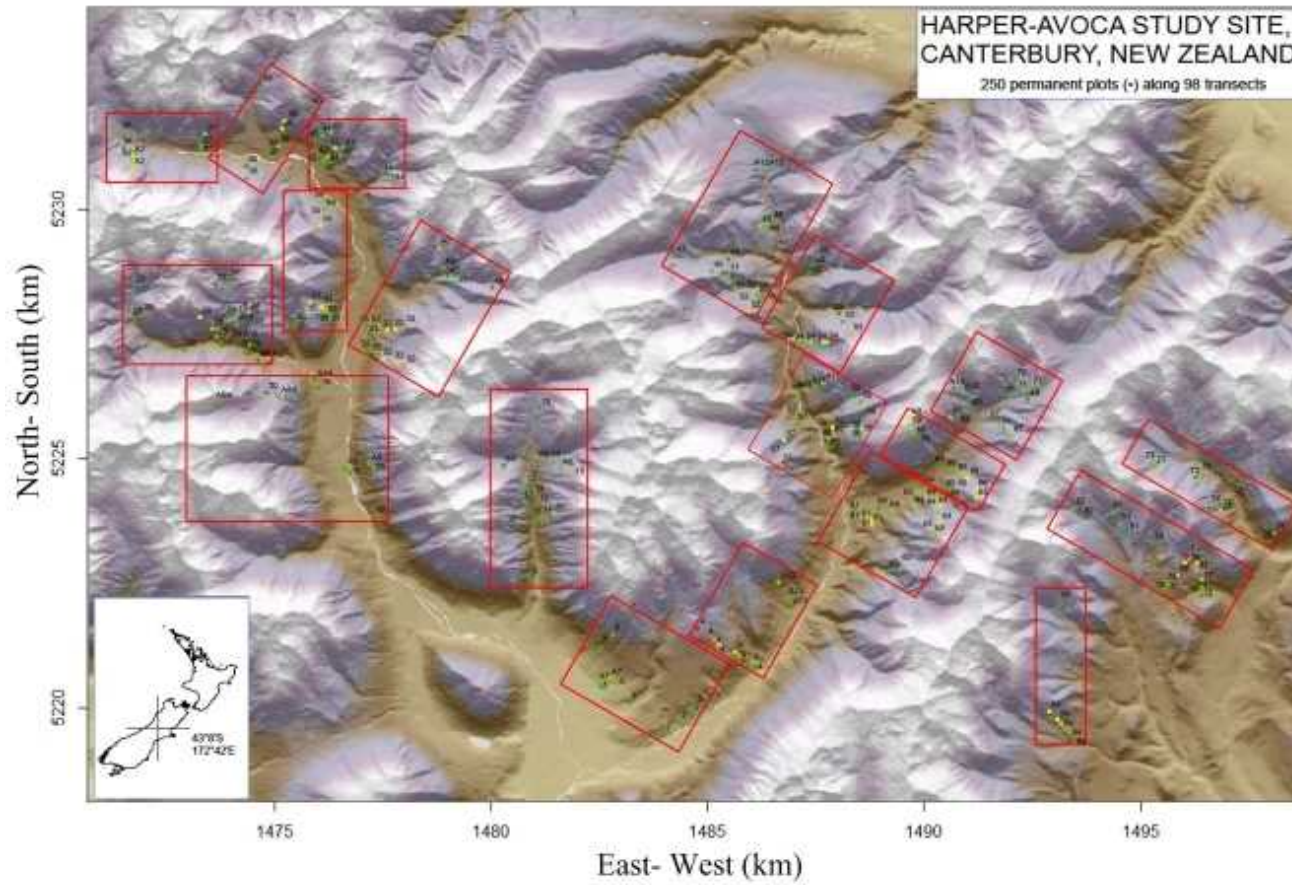
#该图显示每个对象在每种分类组数下的归属（图上每行代表一种组数）。图

#内的表格有不同的颜色，每行两种颜色，代表分两组k=2，三种颜色代表k=3，

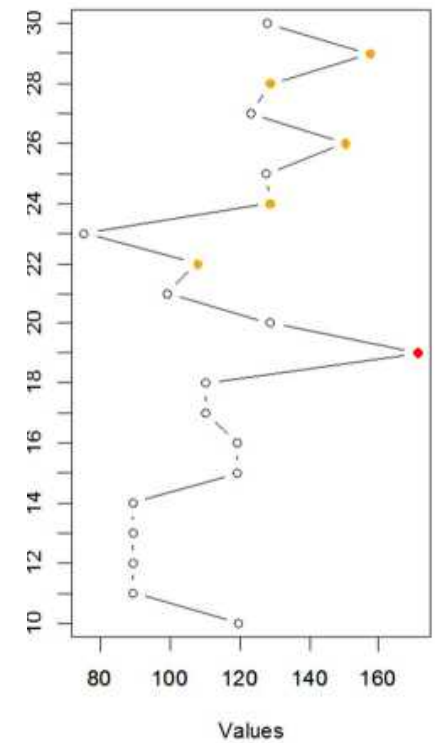
#依此类推。右图代表不同k值条件下的中止标准的统计量。此系列中，到底

#多少组数是最佳方案？如果倾向于较大的组数，哪个是最佳方案呢？

19 blocks



ssi
criterion



多元回归树 (MRT)

物种

<i>Objects</i>	y_1	y_2	y_3	...	y_j	...	y_p
x_1	y_{11}	y_{12}	y_{13}	...	y_{1j}	...	y_{1p}
x_2	y_{21}	y_{22}	y_{23}	...	y_{2j}	...	y_{2p}
x_3	y_{31}	y_{32}	y_{33}	...	y_{3j}	...	y_{3p}
...	...	...	...	...	...	...	...
...	...	...	...	...	...	...	...
...	...	...	...	...	...	...	...
x_i	y_{i1}	y_{i2}	y_{i3}	...	y_{ij}	...	y_{ip}
...	...	...	...	...	...	...	...
...	...	...	...	...	...	...	...
...	...	...	...	...	...	...	...
x_n	y_{n1}	y_{n2}	y_{n3}	...	y_{nj}	...	y_{np}

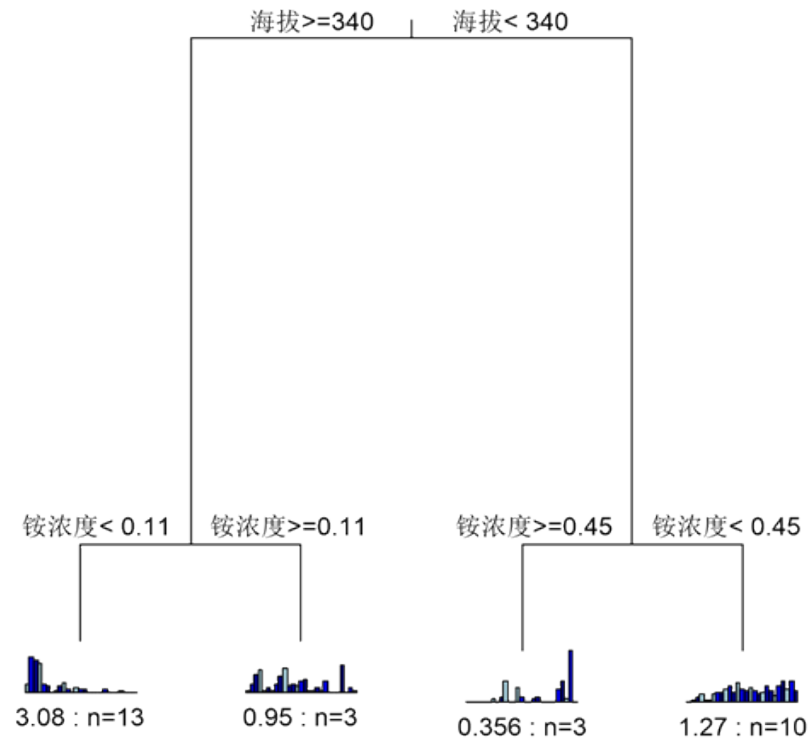
环境因子

<i>Objects</i>	y_1	y_2	y_3	...	y_j	...	y_p
x_1	y_{11}	y_{12}	y_{13}	...	y_{1j}	...	y_{1p}
x_2	y_{21}	y_{22}	y_{23}	...	y_{2j}	...	y_{2p}
x_3	y_{31}	y_{32}	y_{33}	...	y_{3j}	...	y_{3p}
...	...	...	...	...	...	...	...
...	...	...	...	...	...	...	...
...	...	...	...	...	...	...	...
x_i	y_{i1}	y_{i2}	y_{i3}	...	y_{ij}	...	y_{ip}
...	...	...	...	...	...	...	...
...	...	...	...	...	...	...	...
...	...	...	...	...	...	...	...
x_n	y_{n1}	y_{n2}	y_{n3}	...	y_{nj}	...	y_{np}

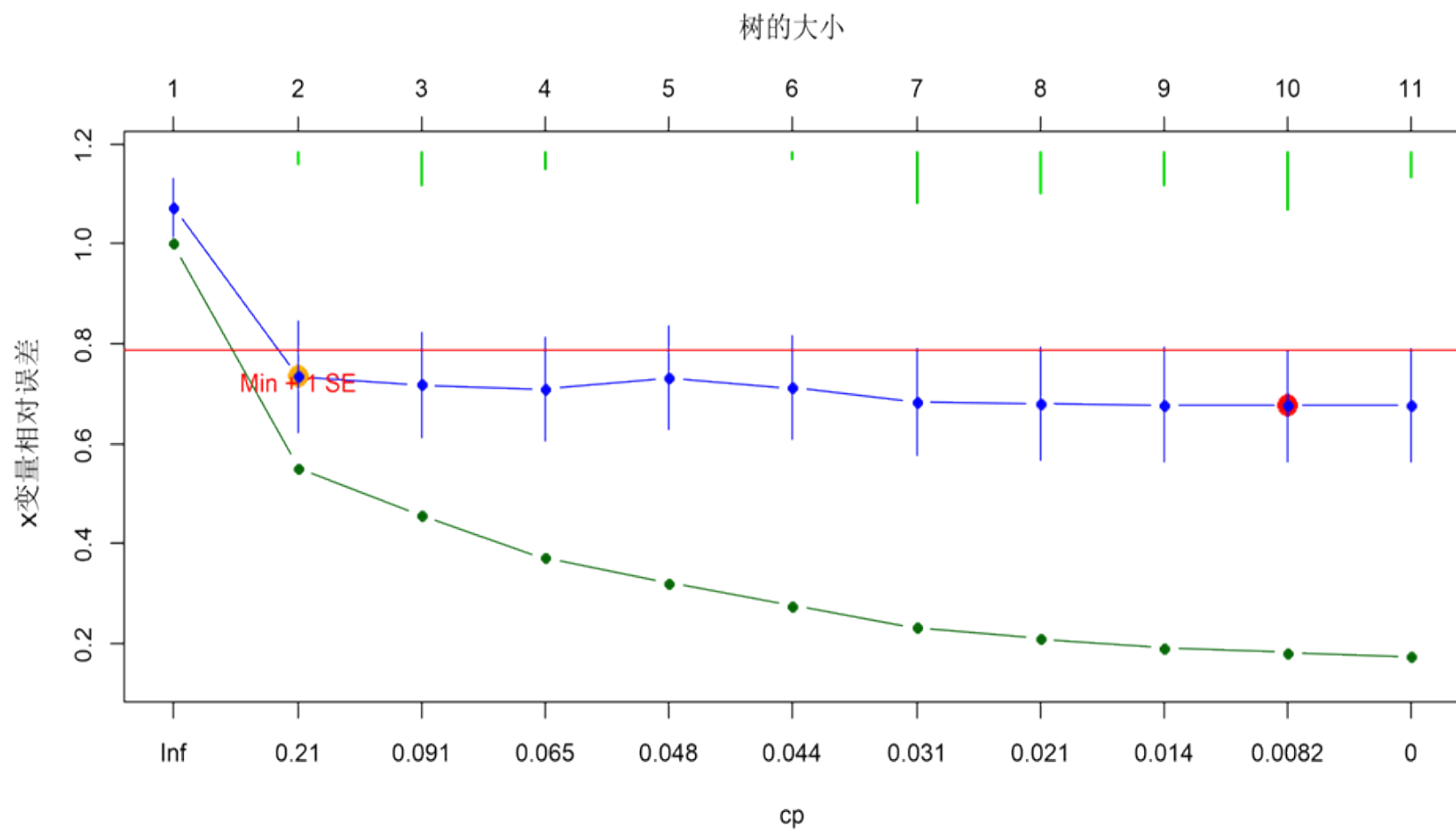
空间

时间

MRT



误差 : 0.37 CV误差 : 0.7 标准差 : 0.101



#多元回归树

library(mvpart) # 载入mvpart程序包

spe.ch.mvpart <- mvpart(data.matrix(spe.norm) ~ ., env, margin=0.08, cp=0,

xv="pick", xval=nrow(spe), xvmult=100, which=4)

#: 参数xv="pick"pick允许通过人机交互的方式选择所需的分类组数。如果想自动选择具

#有最小CVRE值的回归树，可以设定xv="min"。

此处可以点击想要的分组数（例如4组）

MRT作为空间和时间系列约束聚类方法

```
spe.ch.seq <- mvpart(as.matrix(spe) ~ das, env, cp=0, xv="pick", margin=0.08,  
  xval=nrow(spe), xvmult=100, which=4)
```

此时可以点击所期望的分组组数

```
summary(spe.ch.seq)
```

组的组成（终端节点的标识）

```
(gr <- spe.ch.seq$where)
```

重新编排聚类簇的编号

```
aa <- 1
```

```
gr2 <- rep(1,length(gr))
```

```
for (i in 2:length(gr)) {
```

```
  if (gr[i]!=gr[i-1]) aa <- aa+1
```

```
  gr2[i] <- aa
```

```
}
```

在Doubs河地图上标注样方所属的聚类簇

```
plot(spa, asp=1, type="n", main="MRT groups",  
      xlab="x coordinate (km)", ylab="y coordinate (km)")  
lines(spa, col="light blue")  
text(50, 10, "Upstream", cex=1.2, col="red")  
text(25, 115, "Downstream", cex=1.2, col="red")  
k <- length(levels(factor(gr2)))  
for (i in 1:k) {  
  points(spa[gr2==i,1], spa[gr2==i,2], pch=i+20, cex=3, col=i+1, bg=i+1)  
}  
text(spa, row.names(spa), cex=0.8, col="white", font=2)  
legend("bottomright", paste("Group",1:k), pch=(1:k)+20, col=2:(k+1),  
      pt.bg=2:(k+1), pt.cex=2, bty="n")
```

Partitioning beta diversity in a subtropical broad-leaved forest of China

PIERRE LEGENDRE,^{1,6} XIANGCHENG MI,² HAIBAO REN,² KEPING MA,² MINGJIAN YU,³ I-FANG SUN,⁴
AND FANGLIANG HE⁵

¹*Département de Sciences Biologiques, Université de Montréal, C.P. 6128, Succursale Centre-ville,
Montréal, Québec H3C 3J7 Canada*

²*Institute of Botany, Chinese Academy of Sciences, Beijing*

³*College of Life Sciences, Zhejiang University, Hangzhou*

⁴*Center for Tropical Ecology and Biodiversity, Tunghai University, Taichung*

⁵*Department of Renewable Resources, 751 General Services Building, University of Alberta, Edmonton, Alberta T6G 2H1 Canada*

Abstract. The classical environmental control model assumes that species distribution is determined by the spatial variation of underlying habitat conditions. This niche-based model has recently been challenged by the neutral theory of biodiversity which assumes that ecological drift is a key process regulating species coexistence. Understanding the mechanisms that maintain biodiversity in communities critically depends on our ability to decompose the variation of diversity into the contributions of different processes affecting it. Here we

Map of multivariate regression tree classification, 5 groups

2	2	2	1	3	3	3	3	5	2	3	3	3	3	3	3	3	4	4	4	4	4	4	4	3	4	3	5	3	1	1	
2	2	1	2	5	5	5	5	1	1	5	5	3	3	3	3	3	5	5	3	5	5	5	3	3	5	5	5	2	1	1	
2	1	1	2	1	5	5	5	1	1	3	3	3	3	3	3	5	5	3	3	5	5	5	3	3	3	3	1	2	1	2	
1	1	1	1	2	5	5	2	1	1	1	1	5	5	3	5	5	3	3	5	5	5	5	3	1	1	1	1	1	1	2	
1	1	1	2	2	2	2	2	1	1	2	2	5	2	1	2	2	1	3	3	3	5	5			3	1	1	2	1	2	
2	1	2	2	2	2	2	2	1	1	2	2	2	2	1	1	1	1	5	3	3	5	5	5	5	2	2	2	2	2	2	
2	1	2	2	1	1	2	2	1	1	2	2	2	2	1	1	1	1	2	2	5	5	2	2	2	2	2	2	2	1	2	
2	1	1	2	1	2	2	2	1	1	2	2	2	2	1	1	1	1	1	2	2	2	1	1	2	2	2	2	2	1	2	
2	1	1	1	1	2	2	2	1	1	1	2	2	2	1	1	1	1	1	1	2	2	1	1	2	2	2	2	2	1	2	
1	1	1	1	2	2	2	2	1	1	1	2	2	1	1	1	1	1	1	2	2	2	1	1	1	1	2	2	1	2	2	
1	1	1	1	2	2	2	2	2	1	1	1	1	1	1	1	2	2	2	2	2	2	2	2	2	2	2	1	1	1	2	
1	1	1	2	2	1	1	2	2	2	2	1	1	1	1	1	2	2	2	2	2	2	2	2	2	2	2	2	1	1	2	
1	1	1	2	2	1	1	2	2	2	2	1	1	1	1	2	2	2	2	1	1	2	2	2	2	2	2	2	1	1	1	2
2	1	1	1	1	1	1	2	2	2	2	2	1	1	2	2	1	1	1	1	2	2	2	1	1	2	2	2	1	2	2	
2	1	1	1	1	1	2	2	2	2	2	2	1	1	1	1	1	1	2	2	2	2	2	1	1	1	1	2	2	1	2	
2	1	1	1	1	2	2	2	2	1	2	2	1	1	1	1	1	2	2	2	2	2	1	1	2	1	1	2	2	1	2	
2	1	1	1	2	2	2	2	2	1	1	1	1	1	1	1	1	2	2	2	2	2	2	1	1	1	1	1	2	2	1	2
2	2	1	1	2	2	2	2	1	1	1	1	1	1	2	2	2	2	2	2	2	2	2	1	1	1	2	2	2	1	2	
1	2	1	1	1	2	2	1	1	1	1	1	2	2	2	2	2	2	2	1	1	2	2	1	1	1	1	2	2	1	2	
1	1	2	2	2	2	2	2	2	1	2	2	2	1	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2

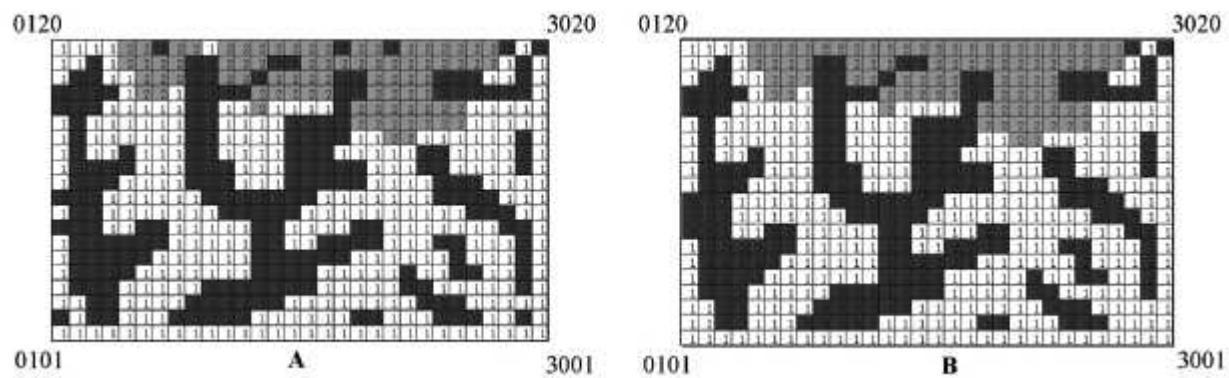
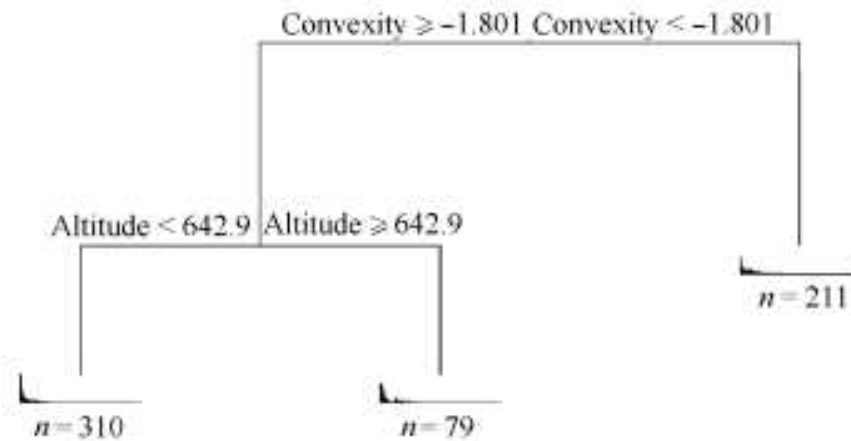
基于多元回归树的常绿阔叶林群丛数量分类——以古田山24公顷森林样地为例

赖江山 米湘成 任海保 马克平*

中国科学院植物研究所植被与环境变化国家重点实验室, 北京 100093

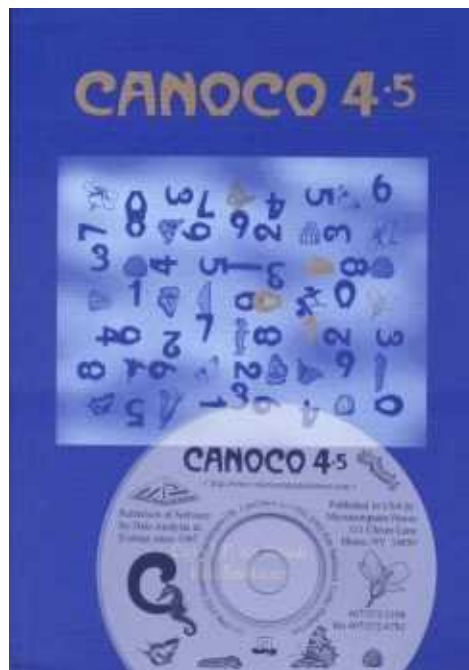
摘 要 常绿阔叶林植被分类一直是植被生态学研究中的难题, 尤其是基本单位——群丛的分类。该文以地形因子和物种组成数据为变量的多元回归树方法, 将浙江古田山24 hm²森林监测样地的森林群落分为3个群丛。所得群丛既反映了群落在时间和空间上的相对间断分布, 也符合植被分类基本单位的特点, 为常绿阔叶林的群丛分类提供了新思路。并首次引入了指示值算法, 解决了以往指示种无法量化的难题。最终以优势层优势种为主、下木层指示种为辅的命名原则, 将3个群丛命名为: 1) 石斑木(*Raphiolepis indica*) + 柳叶蜡梅(*Chimonanthus salicifolius*)—格药枰(*Eurya muricata*) + 赤楠(*Syzygium buxifolium*)—木荷(*Schima superba*) + 甜槠(*Castanopsis eyrei*)群丛; 2) 映山红(*Rhododendron simsii*) + 满山红(*Rhododendron mariesii*)—短柄枹(*Quercus serrata* var. *brevipetiolata*) + 灰白蜡瓣花(*Corylopsis glandulifera* var. *hypoglaucula*)—马尾松(*Pinus massoniana*) + 甜槠群丛; 3) 短尾越橘(*Vaccinium carlesii*) + 毛花连蕊茶(*Camellia fraterna*)—杨梅叶蚊母树(*Distylium myricoides*)—浙江新木姜子(*Neolitsea aurata*)—木荷+甜槠群丛。

关键词 生境, 指示种, 指示值, 地形因子, 植被分类



排序 (Ordination)

- 非约束排序 (Unconstrained Ordination)
- 典范排序 (Canonical Ordination)



Ter Braak



Petr

My Blog

janson的博客

<http://www.sciencenet.cn/blog-laijiangshan.htm>

个人档案



赖江山的博
中国科学院植物研究所
与生物安全研究组助理
加为好友 发

日志

标题：基于CANOCO的生态学数据的多元统计分析（第十章）

浏览：1278 评论：8

基于CANOCO的生态学数据的多元统计分析 10 排序图解读 Visualizing multivariate data第十章已经发布，没有时间仔细推敲，如果有翻译错的地方，敬请指正。10 排序图解读 Visualizing multivariate data



普蘭塔——生态学与生物多样性论坛

中国科学院植物研究所 生物多样性与生物安全研究组创建

帮助 搜索 会员列表 团队 收藏夹
个人资料 站内短信 注销 [janson]

生物统计学 Biostatistics

版主: [janson](#), [Bosmina](#), [tangyaohua](#)

浏览本论坛的会员: [regret](#)

发表新贴

普蘭塔——生态学与生物多样性论坛首页 -> 生物统计学 Biostatistics

分页 1: 1, 2, 3 ... 28, 29, 30 下一页

观看本版的精华主题
标记所有文章为已读

主题	回复	作者	阅读	最新文章
论坛通告				
论坛通告: 地下生态学QQ群 107692061 [0 分页: 1, 2]	26	lyh0455	19242	2010-8-02 周一, 下午4:27 during1111 →
论坛通告: [新闻]中国自然标本馆图片突破百万大关!!!	13	陈建平	3977	2010-8-01 周日, 上午10:24 liuyang99226 →
置顶				
置顶: 基于CANOCO的生态学数据多元统计分析 (中文版) (不断更新) [0 分页: 1 ... 19, 20, 21]	301	janson	41269	2010-8-10 周二, 下午7:44 fritz_zhu83 →
置顶: Canoco 软件的简易教程 [0 分页: 1, 2, 3, 4]	50	smartren	4092	2010-8-09 周一, 下午7:27 wagaag →

生态学多元数据排序分析软件Canoco 5介绍

赖江山

(中国科学院植物研究所植被与环境变化国家重点实验室, 北京 100093)

摘要: 基于样方单元的生物群落调查多元数据是生物多样性研究中最基本的数据类型之一。排序(ordination)作为多元统计最常用的方法之一, 目的是在可视化的低维空间展示多维数据的结构。Canoco 是数据排序分析最流行的软件之一。Canoco 4.5 自从2002年发布以来, 凭借简单的操作界面和功能齐全的绘图工具, 得到广泛的应用。但随着计算机技术的不断发展和新的排序方法不断出现, Canoco 4.5 已经跟不上生态学研究对于多元数据深入分析的需求。作为Canoco 4.5 的升级版, Canoco 5于2012年10月发布。Canoco 5在Canoco 4.5基础上做了很多改进, 并增加一些新的分析方法(比如PCNM)和完善了绘图功能。本文概述Canoco 5所做的主要改进, 并对有些重要操作步骤进行提示, 供同行参考。

关键词: 方差分解, 相邻矩阵主坐标, 非度量多维尺度分析, 谱系, 功能属性

The introduction of Canoco 5: A new version of ecological multivariate data ordination program

Jiangshan Lai

State Key Laboratory of Vegetation and Environmental Change, Institute of Botany, Chinese Academy of Sciences, Beijing 100093

Abstract: Multidimensional data on community composition are fundamental materials for biodiversity research. Ordination is one of the most important multivariate statistical methods. The aim of ordination is to visualize multidimensional data structure at the low-dimensional ordination space. Canoco is one of the most popular programs for ordination analysis. Canoco 4.5 was widely used for ordination analysis after it was released in 2002, with a simple user interface and powerful graph tools. However, with the developments of computer and new ordination methods, Canoco 4.5 has been behind to the requirements of ecologists to deeply analyze multivariate data. As the update version of Canoco 4.5, Canoco 5 has been released in Octo-

765 赖江山

生态学多元数据排序分析软件Canoco 5介绍 **Hot!**

2013 Vol. 21 (6): 765-768 [摘要] (1111) [PDF 1403KB] (1042)

需要Canoco5.03软件的可以留下邮箱

分页 1, 2 下一页

发表新帖

发表回复

普融塔 - 生态学与生物多样性论坛首页 -> 生物统计学 Biostatistics

阅读上一个主题 :: 阅读下一个主题

作者

主题

悠游龙吟

D时间: 2014-9-22 周一, 下午 10:57 标题: 需要Canoco5.03软件的可以留下邮箱

回复 引用 删除 置顶 屏蔽 举报

访客

☆☆☆☆☆

性别: ♂

加入时间: 2011/10/18

文章: 22

来自: 浙江

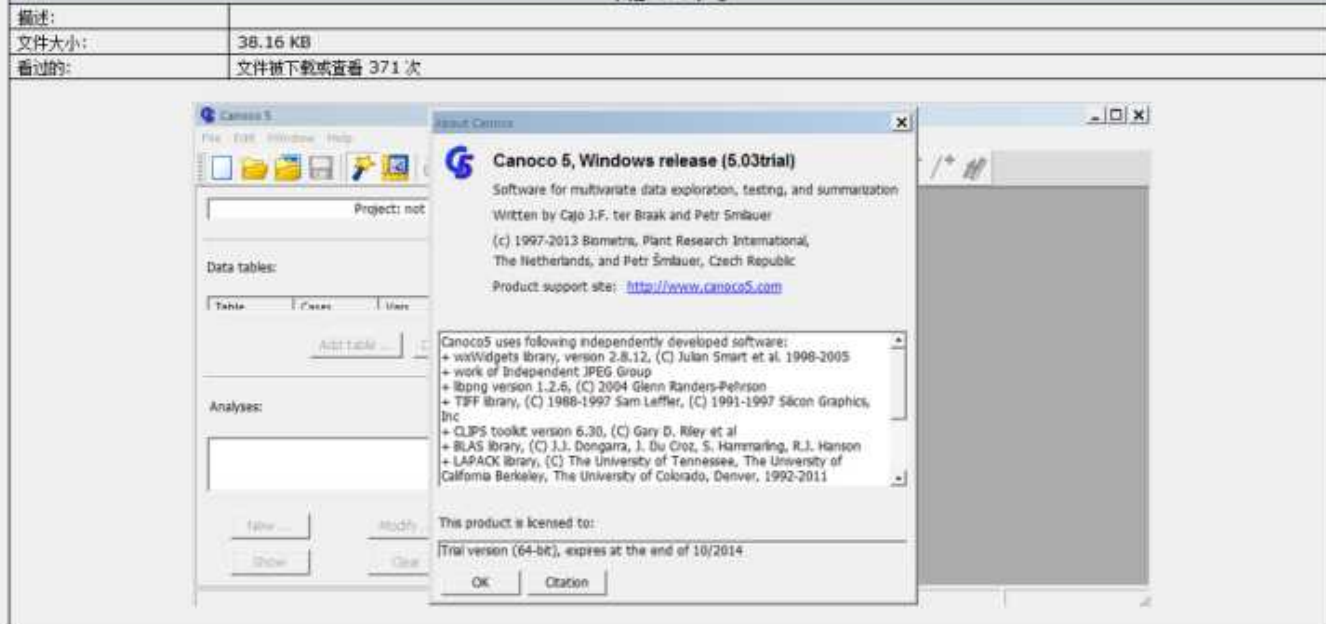
金币: 55.68

[赚与换币]

[修改昵称与头像]

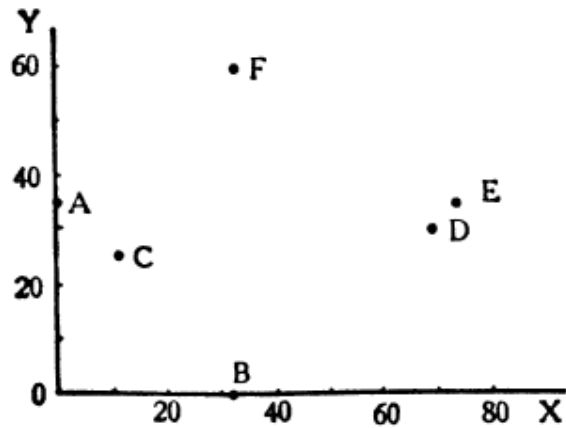
今天很好运地从荷兰外教老师处获得了Canoco5.03软件，经测试可正常安装，在安装时不用许可证License Code，name和company随意填写即可，小用了一下感觉界面改动较大，把原来4版本的功能整合到一起，可以直接从excel文件里导入数据，具体细节有待进一步研究。希望有人下载后不要进一步扩散到网络，毕竟涉及版权，不能太占荷兰人的便宜，我的外教老师知道后估计会很不高兴

图像 478.png

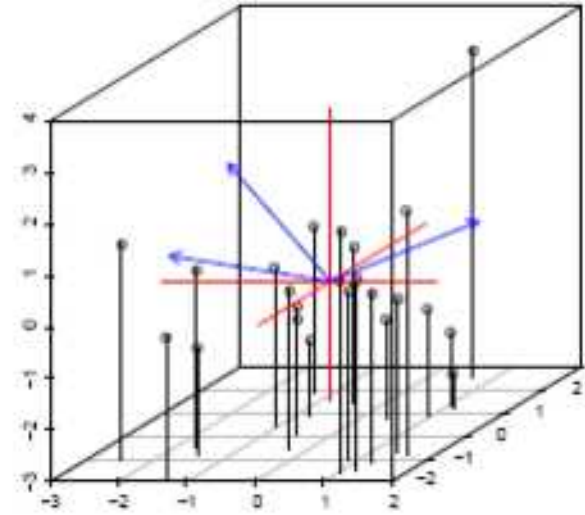


<http://www.planta.cn/forum/viewtopic.php?t=37131&highlight=>

Why ordination?



2个种的排序图

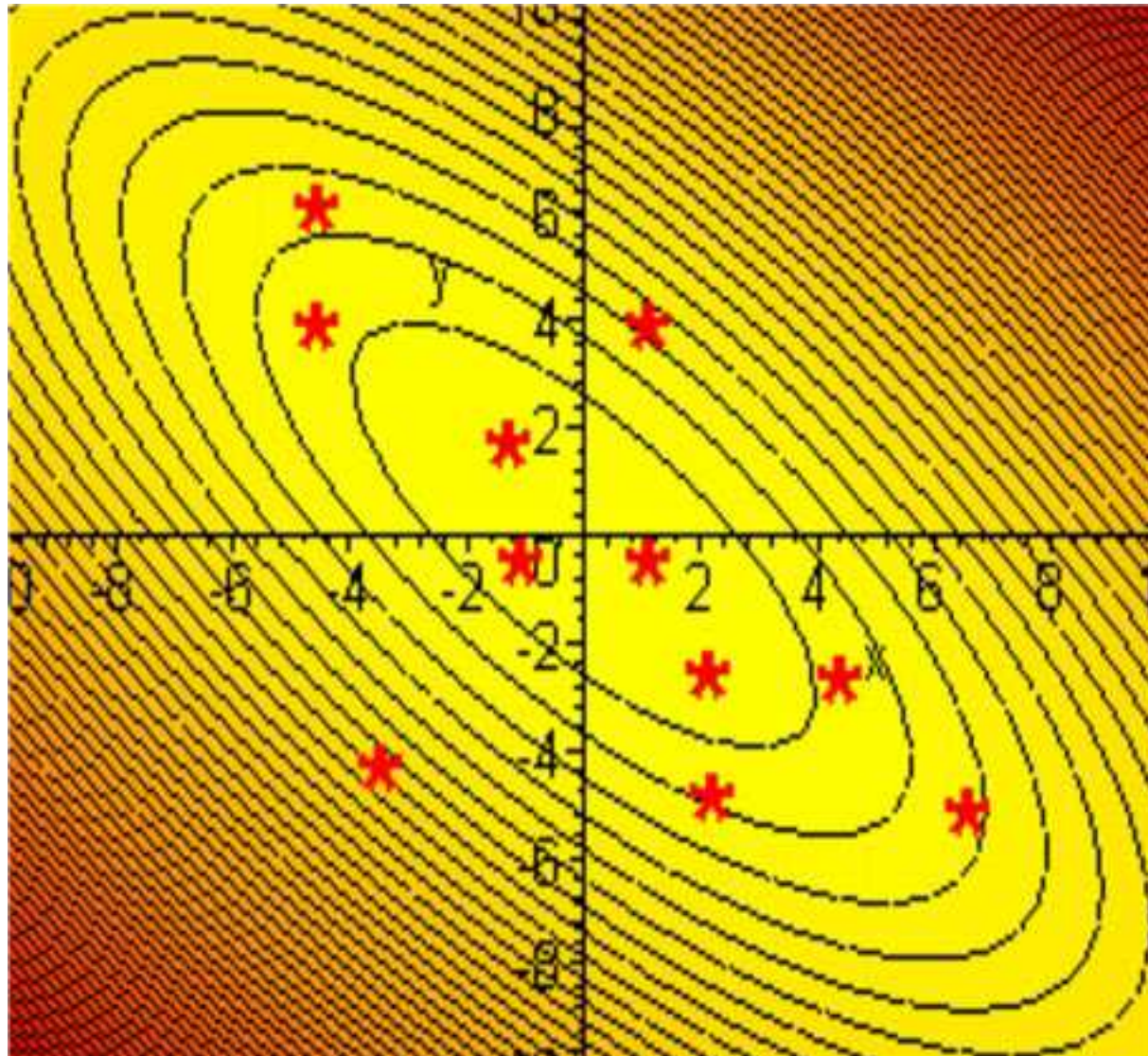


3个种的排序图

4个种的排序图？？？

40个种排序图？？？

Why ordination?



排序的目标：

1.降低维数，减少坐标轴的数目；

2. 由降低维数引起的信息损失尽量少，即发生最小的畸变，也就是让新的坐标系前几轴排序轴包含大量的生态信息。

排序类别

- 非约束排序 **Unconstrained Ordination**

PCA (Principal components analysis)
CA (Correspondence analysis) (DCA)
PCoA (Principal Coordinate Analysis)

} 特征根方法

NMDS(Nonmetric Multidimensional Scaling) 非度量多维尺度分析

- 约束排序(典范排序) **constrained Ordination**

RDA (Redundance analysis)
CCA (Canonical correspondence analysis)
LDA (linear discriminant analysis)

} 非对称

(CCorA), canonical correlation analysis; (CoIA) co-inertia analysis,

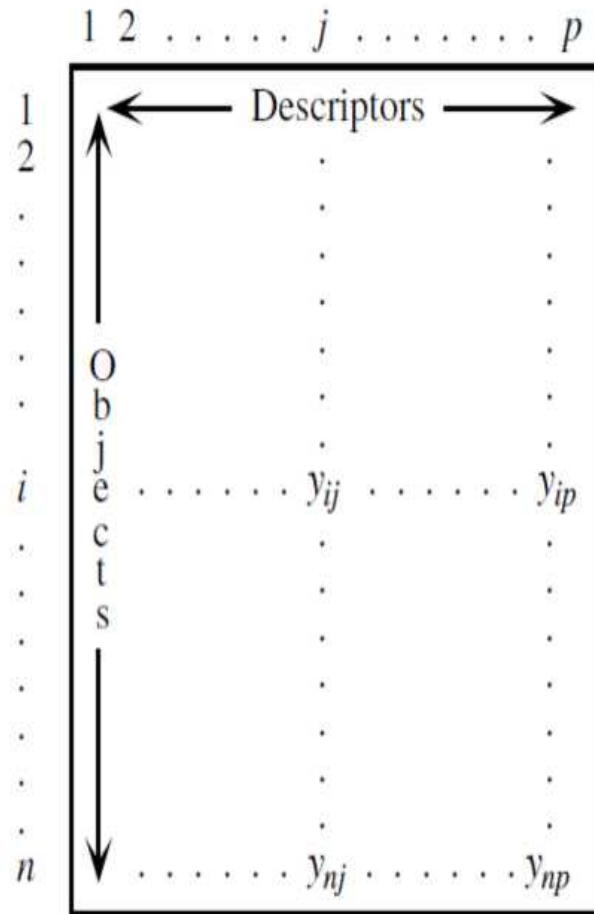
(MFA)multiple factor analysis (对称)

非约束排序

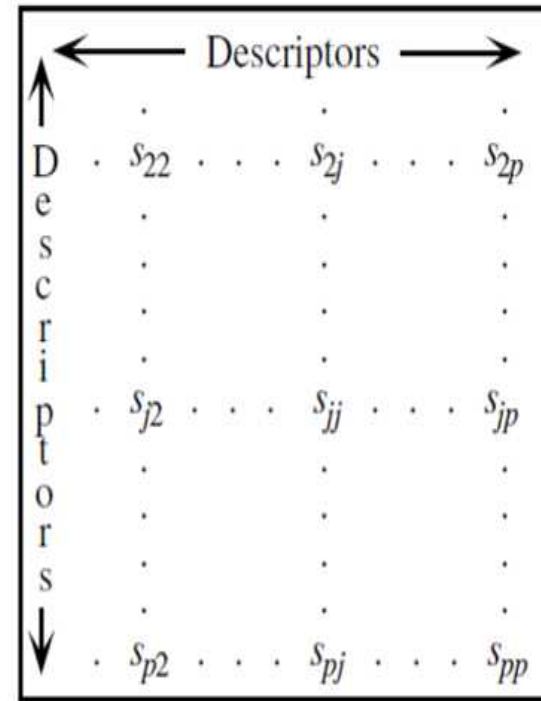
(Unconstrained Ordination)

- 主成分分析（principal component analysis, PCA）：基于特征向量的主要排序方法。分析对象是原始的定量数据。排序图展示样方之间的欧氏距离。
- 对应分析（correspondence analysis, CA）：分析对象必须是频度或类频度、同量纲的非负数据。排序图展示行（对象）或列（变量）之间的卡方距离。在生态学研究主要用于分析物种数据。
- 主坐标分析（principal coordinate analysis, PCoA）：分析对象为距离矩阵（大部分为Q模式），而非原始的样方-变量矩阵表格。因此，可以灵活选择关联测度。
- 非度量多维尺度分析（nonmetric multidimensional scaling, NMDS）：与前面三种排序方法不同，NMDS不是基于特征向量提取的排序方法。NMDS尝试在预先设定数量的排序轴去排序对象，目标是保持这些对象排位关系（ordering relationships）不变。

PCA



Dispersion (covariance) matrix



$$s_{jk} = \frac{1}{n-1} \sum_{i=1}^n (y_{ij} - \bar{y}_j) (y_{ik} - \bar{y}_k)$$

协方差

```
# 载入本章所需的程序包
```

```
library(ade4)
```

```
library(vegan)
```

```
library(gclus)
```

```
library(ape)
```

```
# 载入CSV文件数据
```

```
spe <- read.csv("DoubsSpe.csv", row.names=1)
```

```
env <- read.csv("DoubsEnv.csv", row.names=1)
```

```
spa <- read.csv("DoubsSpa.csv", row.names=1)
```

```
# 删除没有数据的方格
```

```
spe <- spe[-8,]
```

```
env <- env[-8,]
```

```
spa <- spa[-8,]
```

```
# 显示环境量数据集的内容
```

```
summary(env)          # 描述性统计
```

全部环境量数据的PCA分析（基于相关矩阵：参数scale=TURE）

env.pca <- rda(env, scale=T) # 参数scale=TRUE 表示环境量行标准化（不同量不一定）

env.pca

summary(env.pca) # 默认 scaling=2

summary(env.pca, scaling=1)

#注意函数summary（）内的参数scaling，控制排序所出的尺型，

#与函数rda（）内数据标准化的参数scale无关。Scaling 相关内容参考109

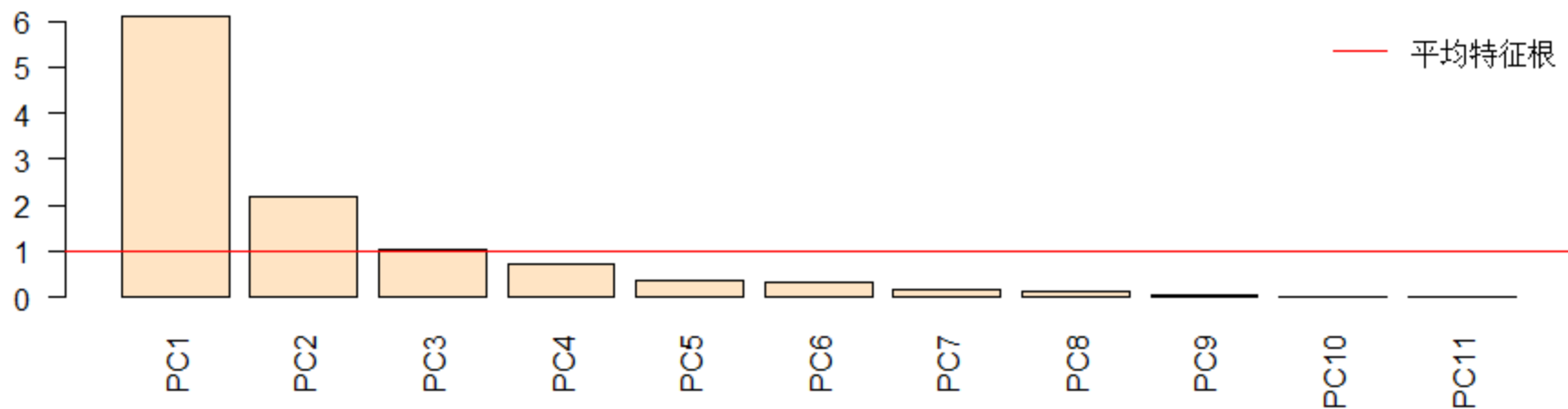
看和控制PCA出的部分果

?cca.object # 解vegan包出的排序果对象结构和如何提取部分果

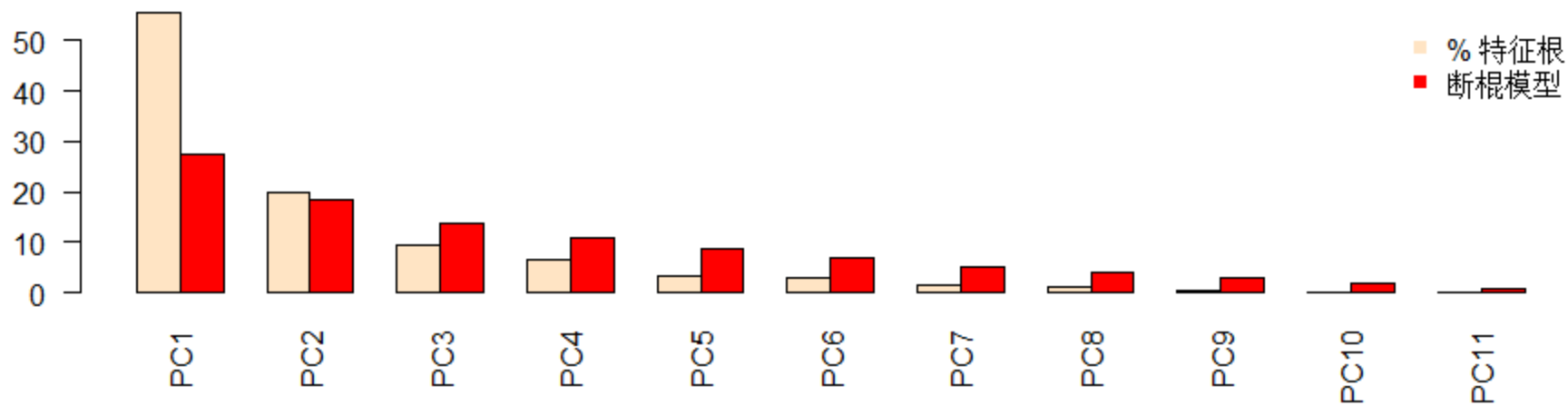
特征根

(ev <- env.pca\$CA\$eig)

特征根



% 变差



```
# 用Kaiser-Guttman准则取排序 （平均） 法
```

```
ev[ev > mean(ev)]
```

```
# 断棍模型 （公式参考：Joseph De Vita 1979 Niche Separation and the Broken-Stick Model, American Naturalist）
```

```
n <- length(ev)
```

```
bsm <- data.frame(j=seq(1:n), p=0)
```

```
bsm$p[1] <- 1/n
```

```
for (i in 2:n) {bsm$p[i] = bsm$p[i-1] + (1/(n + 1 - i))}
```

```
bsm$p <- 100*bsm$p/n
```

```
bsm
```

```
# 制每的特征根和方差百分比
```

```
par(mfrow=c(2,1))
```

```
barplot(ev, main="特征根", col="bisque", las=2)
```

```
abline(h=mean(ev), col="red")      # 特征根平均
```

```
legend("topright", "平均特征根", lwd=1, col=2, bty="n")
```

```
barplot(t(cbind(100*ev/sum(ev), bsm$p[n:1])), beside=TRUE, main="% 差", col=c("bisque",2), las=2)
```

```
legend("topright", c("% 特征根", "断棍模型"), pch=15, col=c("bisque",2), bty="n")
```

以上内容仅为本文档的试下载部分，为可阅读页数的一半内容。如要下载或阅读全文，请访问：<https://d.book118.com/987104162003006133>