

▶ 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

高等教育出版社

» 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

SHULIANG SHENGTAI XUE — R YUYAN DE YINGYONG

高等教育出版社·北京

图字：01 - 2014 - 1754 号

Translation from English language edition:

*Numerical Ecology with R* by Daniel Borcard, François Gillet and Pierre Legendre

Copyright © Springer Science + Business Media, LLC 2011

All Rights Reserved

### 内容简介

近年来,随着新的数据分析方法在生态学和环境科学研究中的迅速发展 and 大数据时代的来临,R 语言统计软件以其灵活、开放、易于掌握、免费等诸多优点,在生态科学和环境科学研究领域迅速传播并赢得广大研究者的青睐。数量生态学方法是现代生态学研究的重要工具,本书是连接数量生态学方法和 R 语言的桥梁。本书首先介绍探索性数据分析和关联矩阵的构建,然后介绍数量生态学的三类主要方法:聚类分析、排序(非约束排序和典范排序)和空间分析。本书的重点不是介绍数量方法的理论基础和数学公式,而是在简要介绍原理的基础上,利用案例数据,手把手地教大家如何在 R 中实现数量分析。本书可作为生态学、环境科学及其他相关领域(例如海洋学、分子生态学、农学和土壤科学)本科生和研究生的教材,也可作为相关专业科研人员的自学参考书。

### 图书在版编目(CIP)数据

数量生态学: R 语言的应用 / (加) 博卡德(Borcard, D.), (法) 吉莱(Gillet, F.) (加) 勒让德(Legendre, P.) 著; 赖江山译. -- 北京: 高等教育出版社, 2014.5  
书名原文: Numerical Ecology with R  
ISBN 978 - 7 - 04 - 039472 - 6

I. ①数… II. ①博… ②吉… ③勒… ④赖… III. ①生态学 - 数量化理论 IV. ①Q14

中国版本图书馆 CIP 数据核字(2014)第 073645 号

策划编辑 柳丽丽 责任编辑 柳丽丽 封面设计 张楠 版式设计 马敬茹  
责任校对 刘娟娟 责任印制 赵义民

出版发行 高等教育出版社  
社址 北京市西城区德外大街 4 号  
邮政编码 100120  
印刷 化学工业出版社印刷厂  
开本 787 mm × 1092 mm 1/16  
印张 18  
字数 330 千字  
购书热线 010 - 58581118

咨询电话 400 - 810 - 0598  
网址 <http://www.hep.edu.cn>  
<http://www.hep.com.cn>  
网上订购 <http://www.landaco.com>  
<http://www.landaco.com.cn>  
版次 2014 年 5 月第 1 版  
印次 2014 年 5 月第 1 次印刷  
定价 49.00 元

本书如有缺页、倒页、脱页等质量问题,请到所购图书销售部门联系调换  
版权所有 侵权必究  
物料号 39472 - 00

# 译者前言

由 Pierre Legendre 和 Louis Legendre 编写的 *Numerical Ecology* 是数量生态学最权威的论著之一。这部论著不仅系统介绍了用于生态学多元数据的统计方法，同时还提供了这些方法的推导过程和数学表达式。通过这部论著，我们不仅可以查阅生态学研究中所需的数量分析方法，也能够了解这些方法背后的数学原理。

大部分生态学研究不仅需要了解数量分析方法的原理和用途，更需要掌握如何实现利用这些方法来分析自己的数据。R 语言由于其具有强大的数据图形展示和统计分析功能，可以免费使用和更新，同时又有大量可随时加载的有针对性的软件包，使得 R 软件成为众多统计分析人员的首选软件。Springer 出版社于 2011 年出版由 Daniel Borcard、François Gillet 和 Pierre Legendre 共同编写的 *Numerical Ecology with R*。该书的特色是在简要介绍分析方法原理的基础上，利用案例数据逐步展示如何在 R 中实现生态学数据数量分析的基本方法（数据描述性统计、关联测度计算）和高级方法（聚类分析、排序分析和空间分析），与 *Numerical Ecology* 互为补充，相得益彰。

2012 年初我看到 *Numerical Ecology with R* 一书后，非常兴奋，随即萌发要把它翻译成中文的念头，目的是让国内同行尽快熟悉如何利用 R 语言实现生态学数据的数量分析，并应用到相关的教学和科研工作中。但翻译权威论著是一项艰难的任务，直到 2013 年初我才下定决心开始着手翻译，历时近一年的时间终于完成。在翻译和校对过程中，我自己也受益匪浅，也是对数量生态学方法再一次的系统学习。

完成这项工作离不开众人的帮助。首先感谢中国科学院植物研究所生物多样性与生物安全研究组七位同事（马克平、米湘成、梁宇、任海保、陈国科、陈磊和刘晓娟）审阅了中文译稿并提出很多宝贵的意见，其次感谢高等教育出版社的李冰祥博士和柳丽丽编辑对此书出版给予的帮助，同时也要感谢付增娟女士在中文稿的审校和文字润色方面提供的帮助。

由于时间和水平有限，翻译过程难免会有疏漏和错误之处，敬请读者批评指正。

赖江山

2013 年 12 月 25 日

于北京香山

# 前言

生态学很迷人，因此，教生态学也是一门艺术，它为有所期待的听众呈现了一个梦幻的主题。当然，教生态学也非常不容易，因为现代生态科学的复杂程度远远超出了高中课程中关于生态学的一般性介绍，也超出了电视中呈现的关于生态系统的奇妙场景。但有所期待的听众已经做好了努力的准备。数量生态学是另外一番景象。不知道什么原因，大部分敬畏自然的人极不情愿量化自然或是用数学工具来理解自然。自然界能够脱离数学吗？显然不可能，因为数学是所有科学共同的语言。因此，作为生物统计学和数量生态学的教师，不得不面对听众的这种不情愿，在开始教他们这门课之前，必须努力说服听众，数量生态学非常有意思，也非常有用。

在过去的几十年中，生态学者，无论是学生还是研究人员（在科研机构、私立机构或政府部门），在收集数据和进行研究时，往往缺乏统计方面的考虑，更愿意委托别人帮他们做数据统计分析。被委托的人基本都是专业的统计学者。的确，很多情况下他们能将统计学的进展及时融合到生态学研究的整个过程。然而，统计学者通常给出的结果仅仅是一些基本的统计和显著性检验方法对数据的总结和归纳，往往不能揭示隐藏在数据背后的丰富的生态学信息。生态学界和统计学界的隔阂产生了很多问题，其中最基本的是生态学家不了解统计方法的运算过程，而统计学家不了解被验证的生态学假说，也不了解生态学数据处理的特殊要求，例如双零问题（double-zero problem）就是一个很好的案例。这种双重不了解不仅阻止了生态学数据的正确分析，也阻止了生态学分析专用统计方法的发展。

或许具有良好数学功底的生态学家能够打破这种局面。幸运的是，过去的四十多年间，这样的生态学家越来越多。他们的努力为数量生态学的发展做出了极大的贡献，有几本优秀的教科书也在此期间出版。同时，生态学家也逐渐意识到研究过程中实验设计和数据分析中统计学的重要性。

然而，长期以来，我们缺乏一套有效且应用广泛的统计软件教学系统。如果没有实践训练，生物统计学和数量生态学的课程便无从谈起。与商业软件联系在一起的生态学课程也不错，但商业软件的使用会有期限限制，过期之后便无法使用或必须重新购买。另外，商业软件往往是为更广泛的群体所设计，一般不会单独为生态学家考虑，所以商业软件并不能包括所有生态学数据

## II 前言

分析的方法。R 语言的出现很好地解决了以上这些问题。我们要感谢许多研究人员制作并无偿贡献出精心设计的通用程序包。现在，老师们不必再说“请大家看书，这是 PCA 的计算过程……”，他们完全可以这么说：“我将在屏幕上给大家演示一下 PCA 的运行过程，无论你在世界的哪个角落，在几分钟内，你们就可以学会如何针对自己的数据运行自己的程序！”

R 语言的另外一个基本特征是它是一个自我学习的软件系统。一本有关 R 语言的书肯定要遵循这个理念，并为所有期望探索自己课题的人提供必要的支持。本书的目的是构建一座连接数量生态学理论和实践的桥梁，我们真诚地希望本书能够造就许多快乐的老师和快乐的生态学家。

加拿大魁北克省 蒙特利尔  
法国 贝桑松  
加拿大魁北克省 蒙特利尔

Daniel Borcard  
François Gillet  
Pierre Legendre

## 郑重声明

高等教育出版社依法对本书享有专有出版权。任何未经许可的复制、销售行为均违反《中华人民共和国著作权法》，其行为人将承担相应的民事责任和行政责任；构成犯罪的，将被依法追究刑事责任。为了维护市场秩序，保护读者的合法权益，避免读者误用盗版书造成不良后果，我社将配合行政执法部门和司法机关对违法犯罪的单位和个人进行严厉打击。社会各界人士如发现上述侵权行为，希望及时举报，本社将奖励举报有功人员。

反盗版举报电话 (010) 58581897 58582371 58581879

反盗版举报传真 (010) 82086060

反盗版举报邮箱 dd@hep.com.cn

通信地址 北京市西城区德外大街4号 高等教育出版社法务部

邮政编码 100120

# 目录

<b>第 1 章 绪论</b> .....	1
1.1 为什么需要数量生态学? .....	1
1.2 为什么用 R? .....	2
1.3 本书的读者群和结构 .....	2
1.4 如何使用本书 .....	3
1.5 数据集 .....	4
1.5.1 Doubs 鱼类数据集 .....	4
1.5.2 甲螨数据集 .....	5
1.6 关于 R 帮助资源的提醒 .....	6
1.7 现在是时候了.....	7
<b>第 2 章 探索性数据分析</b> .....	8
2.1 目标 .....	8
2.2 数据探索 .....	8
2.2.1 数据提取 .....	8
2.2.2 物种数据: 第一次接触 .....	9
2.2.3 物种数据: 进一步分析 .....	10
2.2.4 物种数据转化 .....	16
2.2.5 环境数据 .....	22
2.3 小结 .....	27
<b>第 3 章 关联测度与矩阵</b> .....	28
3.1 目标 .....	28
3.2 关联测度的主要类别 (简短概述) .....	28
3.2.1 Q 模式和 R 模式 .....	29
3.2.2 Q 模式下对称或非对称的系数: 双零问题 .....	29
3.2.3 定性或定量数据的关联测度 .....	30
3.2.4 概括 .....	30
3.3 Q 模式: 计算对象之间的距离矩阵 .....	30
3.3.1 Q 模式: 定量的物种数据 .....	31
3.3.2 Q 模式: 二元 (有-无) 物种数据 .....	33



## II 目录

3.3.3 Q 模式: 定量数据 (除物种多度数据外的数据)	37
3.3.4 Q 模式: 二元数据 (除物种有-无数据外的数据)	39
3.3.5 Q 模式: 混合类型、包括分类 (定性多级) 变量	40
3.4 R 模式: 计算变量之间的依赖矩阵	42
3.4.1 R 模式: 物种多度数据	42
3.4.2 R 模式: 物种有-无数据	43
3.4.3 R 模式: 定量和序数数据 (除物种多度外的数据)	43
3.4.4 R 模式: 二元数据 (除物种多度外的数据)	45
3.5 物种数据的预转化	45
3.6 小结	46
<b>第 4 章 聚类分析</b>	<b>48</b>
4.1 目标	48
4.2 聚类概述	48
4.3 基于连接的层次聚类	50
4.3.1 单连接聚合聚类	50
4.3.2 完全连接聚合聚类	52
4.4 平均聚合聚类	53
4.5 Ward 最小方差聚类	55
4.6 灵活聚类	56
4.7 解读和比较层次聚类结果	56
4.7.1 引言	56
4.7.2 同表型相关	57
4.7.3 寻找可解读的聚类簇	60
4.8 非层次聚类	73
4.8.1 $k$ -均值划分	74
4.8.2 围绕中心点划分 (PAM)	76
4.9 用环境数据进行比较	80
4.9.1 用外部数据进行类型比较 (方差分析途径)	80
4.9.2 双类型比较 (列联表分析)	82
4.10 物种集合	83
4.10.1 组内数据简单统计	83
4.10.2 Kendall 共性系数 ( $W$ )	84
4.10.3 基于有-无数据的物种集合	87
4.10.4 IndVal: 物种指示值	88
4.11 多元回归树: 约束聚类	90

4.11.1	引言	90
4.11.2	计算 (原理)	90
4.11.3	使用 mvpart 和 MVPARTwrap 程序包运行 MRT	92
4.11.4	组合 MRT 和 IndVal	96
4.11.5	作为时序型 (chronological) 聚类方法的 MRT	97
4.12	另类途径: 模糊聚类	99
4.12.1	使用 cluster 程序包内 fanny() 函数进行 c-均值模糊聚类	99
4.13	小结	103
<b>第 5 章</b>	<b>非约束排序</b>	<b>104</b>
5.1	目标	104
5.2	排序概述	104
5.2.1	多维空间	104
5.2.2	降维空间内的排序	105
5.3	主成分分析 (PCA)	106
5.3.1	概述	106
5.3.2	使用 rda() 函数对 Doubs 环境数据进行 PCA 分析	107
5.3.3	转化后的物种数据 PCA 分析	116
5.3.4	PCA 应用领域	117
5.3.5	使用 PCA() 函数进行 PCA 分析	118
5.4	对应分析 (CA)	118
5.4.1	引言	118
5.4.2	使用 vegan 包里的 cca() 函数进行 CA 分析	119
5.4.3	使用 CA() 函数进行对应分析	123
5.4.4	弓形效应和去趋势对应分析 (DCA)	124
5.4.5	多重对应分析 (MCA)	125
5.5	主坐标分析 (PCoA)	125
5.5.1	引言	125
5.5.2	利用 cmdscale 包和 vegan 包对 Doubs 数据进行 PCoA 分析	126
5.5.3	使用 pcoa() 函数对 Doubs 数据进行 PCoA 分析	128
5.6	非度量多维尺度分析 (NMDS)	130
5.6.1	引言	130
5.6.2	鱼类数据 NMDS 分析	131
5.7	手写排序函数	134
<b>第 6 章</b>	<b>典范排序</b>	<b>137</b>
6.1	目标	137

6.2	典范排序概述 .....	137
6.3	冗余分析 (RDA) .....	138
6.3.1	引言 .....	138
6.3.2	Doubs 数据集 RDA 分析 .....	139
6.3.3	手写 RDA 函数 .....	174
6.4	典范对应分析 (CCA) .....	177
6.4.1	引言 .....	177
6.4.2	Doubs 数据集 CCA 分析 .....	178
6.5	线性判别式分析 (LDA) .....	185
6.5.1	引言 .....	185
6.5.2	使用 <code>lda()</code> 函数进行判别式分析 .....	185
6.6	其他非对称分析 .....	188
6.7	两个 (或多个) 数据集的对称分析 .....	188
6.8	典范相关分析 (CCorA) .....	189
6.8.1	引言 .....	189
6.8.2	使用 <code>CCorA</code> 函数进行典范相关分析 .....	189
6.9	协惯量分析 (CoIA) .....	191
6.9.1	引言 .....	191
6.9.2	使用 <code>ade4</code> 包进行协惯量分析 .....	192
6.10	多元因子分析 (MFA) .....	195
6.10.1	引言 .....	195
6.10.2	使用 <code>FactoMineR</code> 进行多元因子分析 .....	196
6.11	小结 .....	200
<b>第 7 章</b>	<b>生态学数据空间分析</b> .....	<b>202</b>
7.1	目标 .....	202
7.2	空间结构和空间分析: 简短概述 .....	202
7.2.1	引言 .....	202
7.2.2	诱导性空间依赖和空间自相关 .....	203
7.2.3	空间尺度 .....	204
7.2.4	空间异质性 .....	205
7.2.5	空间相关或自相关函数和空间相关图 .....	206
7.2.6	空间相关检验的条件 .....	210
7.2.7	模拟空间结构 .....	211
7.3	多元趋势面分析 .....	211
7.3.1	引言 .....	211

7.3.2 练习趋势面分析 .....	212
7.4 基于特征根的空间变量和空间建模 .....	216
7.4.1 引言 .....	216
7.4.2 基于距离的经典 MEM (之前被称为 PCNM) .....	216
7.4.3 更广泛的 MEM: 除地理距离外的权重 .....	234
7.4.4 应该使用正空间相关还是负空间相关? .....	248
7.4.5 具有方向性的非对称特征向量图 (AEM) .....	248
7.5 另外一种了解空间结构的途径: 多尺度排序 .....	254
7.5.1 原理 .....	254
7.5.2 甲螨数据多尺度排序: 探索性方法 .....	254
7.5.3 去趋势甲螨物种和环境数据多尺度排序 .....	257
7.6 小结 .....	260
参考文献 .....	261
索引 .....	270

# 第 1 章 绪 论

## 1.1 为什么需要数量生态学？

虽然生态数据多元统计分析在 20 世纪 60 年代就已经存在并取得积极进展，但真正繁荣是在 20 世纪 70 年代之后，主要表现为在此期间出版了很多教科书，最具代表性的是 Pierre Legendre 和 Louis Legendre 兄弟编写的《数量生态学》（法文版）（Legendre 和 Legendre, 1979）及其英文译本（Legendre 和 Legendre, 1983）。这些书的作者们对应用于生态学的统计方法进行了认真归纳总结，形成了一门新的学科：数量生态学。这些教科书不仅能为研究人员查阅数量分析方法提供方便，也告诉他们如何选择合适的方法实现自己的研究目标。同时，这些书还提供了数量分析方法的数学表达式和推导过程，使读者有机会了解数量方法背后的数学原理，而不仅仅停留在简单使用的层面。

也就从这个时期开始，数量生态学变得非常流行。每个严谨的研究人员已经意识到使用数量方法最大程度去挖掘辛辛苦苦获得的数据是多么重要。从此，其他数量生态学方面的专业书籍也不断出现（例如，Orlóci 和 Kenkel, 1985; Jongman 等, 1995; McCune 和 Grace, 2002; McGarigal 等, 2000; Zuur 等, 2007）。Legendre 和 Legendre 编写的《数量生态学》英文版第二版也在 1998 年出版，补充了很多第一版没有的数量方法，内容也更加丰富。当然，自从 1998 年以来，数量生态学有很多突破，新方法也层出不穷。本书也会提及我们认为的重要进展和有用的新方法，我们描述的角度更倾向于如何在 R 语言中实现新的方法。当然，对于最新的方法，我们也会提供一些基础性的解释，为读者了解这些新方法提供必要的帮助。

本书并没有包含所有的数量生态学方法。除了最常用和最有效的数量方法之外，我们对其他方法的选择就基于我们作为数量群落生态学家的经验。本书除了介绍一些最主要的方法之外，对涉及的一些非主流的方法也进行了简单的描述。

## 1.2 为什么用 R?

近几年，R 语言的发展和用户增加的速度让人惊叹不已。虽然数量生态学的运算并不是非用 R 语言不可，但不可否认的是，生态学不同领域越来越多的方法都可以用 R 计算，毫不夸张地说，某些最新的方法也只能通过这种自由的 R 程序包才能得以迅速传播。

对于什么是 R 语言本书不做详细介绍，读者可以到 CRAN (the Comprehensive R Archive Network) 网站 (<http://www.R-project.org>) 获得有关 R 语言的知识，网站里有很多链接可以下载各种免费学习 R 的参考资料。对于阅读本书的读者，我们希望你们能够掌握基本的 R 语言操作及语法，例如如何将数据导入 R，对象（向量、矩阵、数据框和因子）生成、调取、处理等。第 2 章的内容将会涵盖一部分 R 的基本操作以及多元数据矩阵运算，整本书一直会使用多元数据矩阵作为分析对象。但很多使用者对多元数据矩阵的运算并不十分熟悉。

本书的目标不在于尽可能罗列某种数量分析的所有 R 函数。对于某种分析方法，我们会运用一个或几个函数，但 R 中也存在其他具有相同功能的函数。本书通常选择一些功能比较完善的程序包中的函数，同时也会用一些自编函数，必要时会对某些程序包做解释说明。需要声明的是我们使用的函数和程序包并不意味着优于未被用到的函数和程序包。

## 1.3 本书的读者群和结构

本书适用于具有一定的一般统计及多元统计分析基础并希望利用 R 语言分析数据的科研人员、研究生和教师，以及希望将他们学科上的知识应用到实践中的朋友。本书大纲组织结构和参考文献与《数量生态学》第二版 (Legendre 和 Legendre, 1998) 基本相同，目的在于方便读者查阅许多数量方法的原理。

本书以应用为导向，内容简单、通俗易懂。每章开头部分简单总结该章的主要内容，以方便读者了解本章的范围，理解本章介绍的分析方法观点。每章引言部分的繁简程度可能不同，主要依据该内容是否在其他统计学教科书中出现。

总之，本书以生态学家的角度引导读者如何运用 R 语言实现主要的多元

统计分析方法。第2章是探索性数据分析，第3章是大多数技术的关键模块的构造，即多元数据矩阵计算和关联分析，接下来的三章是多元统计的三类主要方法：聚类分析（第4章）、非约束排序和典范排序（第5、6章），最后是空间分析（第7章）。本书从数据描述、探索到预测，涵盖了目前生态学数据多元统计分析的基本方法，目的在于为广大读者提供模板式的工具书。

## 1.4 如何使用本书

使用本书电脑操作必不可少，希望读者阅读此书的同时能在电脑上运行 R 代码。为了达到更好的效果，建议读者按顺序阅读各章节，因为每章都以前面的章节为基础。每章学习时务必将空白的 R 工作空间打开。本书所用的数据、R 代码脚本以及 CRAN 网站没有提供但本书使用的函数和程序包都可以从 Springer 出版社网站下载（<http://www.springer.com/978-1-4419-7975-9>）<sup>①</sup>。有些替代 R 函数的自编函数（例如某些控制图像输出的函数）也已经融入目前的 R 代码脚本中。

每章的所有代码（除了极少部分人机交互的代码）都可以通过拷贝-粘贴一次性在 R 工作空间运行，为了能够仔细了解每个函数的功能，我们建议读者还是逐步运行更好。虽然很多函数都通过终止符“#”直接在代码后面加注解，但我们还是希望读者能够多使用 R 自带的帮助系统（在 R 工作空间内输入？及函数名即可）。本书没有详尽解释每个函数所有的参数选项，这项工作不必要也没法完成。相信很多好学的读者肯定不满足本书所提供的案例，会不断探索书中所列函数参数的其他选项，我们也鼓励这样的探索。

每章简短的引言之后，读者可以进入实际案例的练习。每章 R 代码开头部分已经载入本章将使用的 R 程序包和对象，因此，每章的 R 代码都是独立完整的，前后章节无包含关系。

在 R 的日常应用中，几乎很少将每个运算结果都赋予对象名或每个图表都新开一个窗口。但本书尽可能这么做，目的在于方便读者对有些运算的结果进行比较。特别是有些费时的运算过程（例如随机置换检验运算），如果不及时将结果存储为对象，后来需要调用时再重新运行往往非常费时，特别是对于速度慢的计算机，这样做更是得不偿失。

---

<sup>①</sup> 也可以从著者为本书英文版设立的官网 <http://adn.biol.umontreal.ca/~numerical ecology/numecolR/> 下载数据和不断更新的代码脚本。本书所用带中文解释的脚本可以从译者博客 <http://blog.sciencenet.cn/u/laijiangshan> 下载。——译者注

在本书代码部分，所有生成新的绘图窗口的代码已经删除，但从网站下载的是完整的电子版代码。本书也只是选择显示一部分图，而不是全部的图，有些图在 R 里是彩色的，但书中只有黑白印刷版，目的是激励读者务必真正在 R 中运行这些代码，以获得漂亮的图像。

有些代码之后灰色框内的内容是告诉读者这些代码特别的功能和使用窍门。

虽然本书提供了很多研究案例，但并没有提供所有案例分析结果的生态学解释。因为对于相同的分析结果，可能会有不同的生态学解释，所以将结果留给读者自己判读未尝不是件好事。本书的主要目的还在于使读者了解如何通过 R 获得所需要的分析结果和图像。

最后，对于熟悉 R 语言或计算机语言的读者，本书鼓励自己编写函数代码。因此，每章最后的部分都有“自写代码角”，给出计算公式，让读者自己编写函数，并引导读者与 R 自带的函数进行比较。当然，这部分的目的不在于考察读者的计算机能力，而是方便于教学。这些自写函数的案例主要从《数量生态学》(Legendre 和 Legendre, 1998) 中抽取出来，大部分涉及矩阵的代数运算，这些运算在相关程序包都能找到对应的函数。总体来说，自写函数的目的在于使读者更深入地理解一些重要分析方法的数学运算过程。

## 1.5 数据集

除了极个别的案例需要设定一些特定的数据进行演示之外，本书所有的案例基本都以两个主要的数据集作为基础，这两个数据集是 R 程序包自带的数据集。由于 R 的程序包经常更新，数据集也可能更换。因此，本书也以电子表格的形式提供数据集，以保证读者在练习过程中输出的结果与本书一致。本章简要介绍两个数据，第一个 Doubs 数据集在第 2 章中会更详细地进行探索性分析。我们也鼓励读者运用第 2 章介绍的探索性方法分析第二个数据集。

### 1.5.1 Doubs 鱼类数据集

Verneaux (1973, 也见 Verneaux 等, 2003) 在他的博士学位论文中建议使用鱼的种类划分欧洲河流的生态区。Verneaux 认为鱼类群落对水体质量有很好的生物指示作用。Verneaux 以 Doubs 河为例，从源头开始，提出将 Doubs 河分为 4 个以指示鱼类命名的区域：鳟鱼区（指示鱼类——褐鳟 *Salmo trutta fario*）、茴鱼区（指示鱼类——茴鱼 *Thymallus thymallus*）、触须鱼区（指示鱼类——鲃鱼 *Barbus barbus*）和鳊鱼区（指示鱼类——欧鳊 *Abramis brama*）。前



两个区域又可以归为“鲑鱼区域”，后两个区域归为“鲤鱼区域”。4个区域按顺序代表水质由好变坏，从相对干净、养分贫瘠的富氧区域到富营养化的无氧区域变化。

本书所用的 Doubs 数据集是 Verneaux 在他的研究中所收集数据的一部分，来自沿法国和瑞士边境的 Jura 山脉的 Doubs 河 30 个取样点（以下称为样方）的数据。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（笛卡儿坐标系， $x$  和  $y$ ）。这些数据已经作为很多新数量方法的检验案例（Chessel 等，1994）。

R 中 `ade4` 程序包内附带有 Doubs 环境因子数据（本书编写时 `ade4` 的版本是 1.4-14）。在本书提供的数据集内，我们纠正了 `das` 变量里的小错误，同时恢复这些变量的原始单位（表 1.1）。

表 1.1 本书所用 Doubs 数据集环境变量名称及单位

变量名称	代码	单位 <sup>①</sup>
离源头距离	das	km
海拔	alt	m a. s. l.
坡度	pen	%
平均最小流量	deb	$\text{m}^3 \cdot \text{s}^{-1}$
pH	pH	-
钙浓度（硬度）	dur	$\text{mg} \cdot \text{L}^{-1}$
磷酸盐浓度	pho	$\text{mg} \cdot \text{L}^{-1}$
硝酸盐浓度	nit	$\text{mg} \cdot \text{L}^{-1}$
铵浓度	amm	$\text{mg} \cdot \text{L}^{-1}$
氧含量	oxy	$\text{mg} \cdot \text{L}^{-1}$
生物需氧量	dbo	$\text{mg} \cdot \text{L}^{-1}$

### 1.5.2 甲螨数据集

甲螨（蟬螨目：甲螨亚目）是一类个体非常小（0.2 ~ 1.2 mm）、但种类繁多的植食性和腐殖性土壤节肢动物。在透气性良好的土壤或复杂基质上，例

<sup>①</sup> 本书中案例的数据、图表中存在量和单位用法不符合现行国标之处，由于是软件应用类译著，无法修改，均保留原用法，特此说明。——译者注