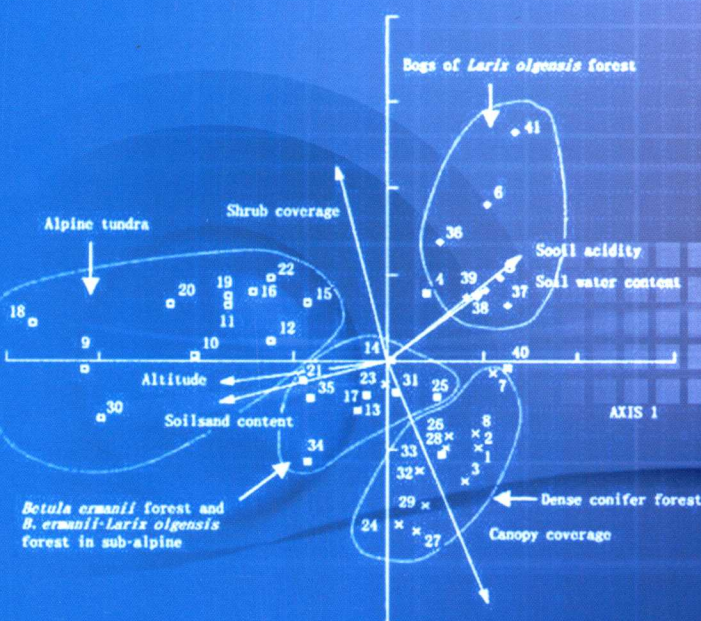




生态学数据分析

——方法、程序与软件

ECOLOGICAL DATA ANALYSES
— METHODS, PROGRAMS AND SOFTWARE

郭水良 于 晶 陈国奇 编 著



科学出版社

生态学数据分析

——方法、程序与软件

郭水良 于 晶 陈国奇 编著



科学出版社

北 京

内 容 简 介

生态学研究需要从原始数据出发,通过系列计算分析,最后作出具有生态意义的解释。本书主要介绍 PCORD for Windows、Canoco for Windows、Curve Expert、SPSS 和 PAST 等软件在生态学数据转换、标准化、函数拟合、遗传多样性、物种多样性、生态位、空间分布格局、聚类、排序和实验数据统计检验上的应用方法。应用 QBASIC 和 R 语言对书中的部分方法编制程序。书中的每一种方法均提供模拟数据,介绍具体的操作步骤。

本书可作为高等院校生态学、环境科学、水生生物学、植物学及其相关专业的研究生教材,也可以作为上述领域的科研和技术人员的工具书。

图书在版编目(CIP)数据

生态学数据分析:方法、程序与软件 / 郭水良,于晶,陈国奇编著. —北京:科学出版社,2015.4
ISBN 978-7-03-043928-4

I. ①生… II. ①郭… ②于… ③陈… III. ①生态学—数据处理 IV. ①Q141

中国版本图书馆 CIP 数据核字(2015)第 055666 号

责任编辑:陈 露 纪四穗
责任印制:谭宏宇/封面设计:殷 靓

科学出版社出版

北京东黄城根北街 16 号

邮政编码:100717

<http://www.sciencep.com>

南京展望文化发展有限公司排版

江苏省句容市排印厂印刷

科学出版社出版 各地新华书店经销

*

2015 年 4 月第 一 版 开本:787×1092 1/16

2015 年 4 月第一次印刷 印张:17

字数:379 000

定价:68.00 元

Proface

前言

在生态学研究,往往涉及大量生物和环境因子的观测数据,不对这些数据进行定量分析,难以寻找出生物种类之间、生物与环境之间的内在联系和生态学规律。数量生态学研究就是从一组原始数据出发,通过一系列计算分析,最后作出具有生态意义的解释。

目前已有一些关于生态学数据处理原理与方法的书籍。例如《植被数量生态学方法》(张金屯,1995)、《数量生态学》(张金屯,2004)、《群落生态学原理与方法》(赵志模等,1990)、*Quantitative Plant Ecology* (Greig-Smith P P, 1983)、《植物生态学数量分类方法》(阳含熙等,1981)、*Quantitative Ecology: Measurement, Models and Scaling*。但是这些书籍主要介绍的是生态学数据分析的原理与方法,并没有提供方法所需要的软件操作步骤和有关计算程序。学生在数据处理时,需要从零散的软件中寻找相应的操作技巧,或者编写相应的计算程序,后者对于多数在生物系的学生来讲有一定难度。2014年,为提高研究生教学质量,上海师范大学开始研究生精品课程建设项目。针对生态学研究在数据处理上遇到的困惑,开展“生态学原理与数据分析”精品课程建设,编写此书。

本书主要介绍群落生态学研究中的数据转换、标准化、数据间关系的函数拟合、生态位、物种和遗传多样性、空间分布格局、聚类、排序、结果图形表达,以及实验生态学中的数据统计检验有关的方法。书中涉及的软件有 PCORD for Windows 4.1、Canoco for Windows、CurveExpert 1.3 和 PAST 2.17 等。同时应用 QBASIC 和 R 语言对书中的一些方法编写程序。每一种方法均提供模拟数据,并进行具体操作步骤的介绍。

对于大多数生态学数据处理方法,本书并没有介绍它们深奥的数学原理和复杂的公式推导,而是将重点放在相关软件、程序的使用方法上,尽量做到图文并茂,清晰明了。每一种方法均提供模拟数据和主要结果的说明。读者只要根据书中的方法,按部就班,就能够完成数据的处理。

对于聚类分析、排序分析的一些常用方法,本书同时介绍 PAST、PCORD、Canoco 软件和 R 语言编写的程序。在实际研究中,读者可以根据手头掌握的软件,选择其中的一种即可完成相应操作。对于生态位宽度和重叠值、种群空间分布格局、图论中的最大生成

树法和多样性指数计算等,本书还编写相应的 QBASIC 程序,读者根据这些程序生成可执行文件,也能够方便地进行数据分析。

生态学研究经常会涉及变量关系的拟合,以构建叶绿素荧光参数的快速光响应曲线、光-光合作用曲线、Logistic 曲线和生物分布与环境关系的高斯曲线等。本书介绍应用 CurveExpert 1.31 进行变量关系拟合的方法。对于实验生态学中的数据,读者需要开展单因素方差分析、双因素方差分析、独立样本 t 检验、成对样本 t 检验等。本书也专门介绍应用 SPSS 软件和 R 语言编写的程序进行此类数据处理的方法。

当代的计算机操作系统、应用软件和编程语言更新迅速。随着时间的推移,书中介绍的一些软件也会更新,或者在新的计算机操作系统上有不兼容的现象,这需要读者及时更新软件版本。

本书由上海师范大学生命与环境科学学院苔藓植物研究组的郭水良、于晶、陈国奇合作编撰。于晶编写第一、二章,郭水良编写第三~十章、十二章,陈国奇编写第十一章。全书由郭水良统稿。

本书的出版得到上海师范大学精品课程建设项目(项目编号: B-6001-13-103414)、上海市科学技术委员会重点项目(项目编号: 12490502700)、上海师范大学内涵建设项目经费的资助。

由于作者水平有限,加之时间仓促,书中难免存在疏漏或不足之处,祈请读者批评指正。

作者

2014 年 11 月

于上海

Contents

目 录

前言

第一章 生态学数据收集	1
第一节 样方设置	1
一、主观取样	1
二、客观取样	1
第二节 样方形状、大小和数目	2
一、样方的形状	2
二、样方的大小	3
三、样方的数目	3
第三节 无样地取样法	4
第二章 群落数量和环境特征	6
第一节 群落的数量特征	6
一、多度	6
二、密度	6
三、盖度	7
四、频度	8
五、高度	8
六、生物量和产量	8
七、优势度和生态重要值	8
八、生活型	8
九、种—面积曲线和群落最小面积	9
第二节 环境特征	9
一、气候数据	9
二、地形数据	10

三、土壤数据	10
四、水体环境	10
五、生物因子数据	11
第三章 数据准备	12
第一节 数据的类型	12
一、名称数据	12
二、顺序性数据	12
三、数量数据	13
四、数据矩阵	13
第二节 数据预处理	14
一、数据简缩	14
二、数据转换	14
三、数据标准化	14
第三节 QBASIC 编程	17
第四章 回归分析	25
第一节 应用 CurveExpert 进行函数拟合	25
一、CurveExpert 软件简介	25
二、操作步骤	25
第二节 应用 SPSS 进行曲线拟合	29
第三节 运用 Excel 电子表格进行变量关系拟合	31
第四节 多元线性回归	31
第五章 多样性指数计算	34
第一节 生物多样性概述	34
第二节 遗传多样性	34
一、取样方法	34
二、遗传多样性标记介绍	35
三、分子数据分析方法	37
四、应用 Popgen32 计算群体遗传多样性和分化	40
第三节 物种多样性	45
一、物种多样性的概念	45
二、 α 物种多样性计算公式	45
三、 β 物种多样性计算公式	48

四、物种多样性的计算程序和软件	48
第六章 种间关系分析	52
第一节 相关分析	52
一、连续数据的相关性分析	52
二、二元数据的双系列相关系数	52
第二节 种间关联	53
第三节 种间相关	54
一、Pearson 相关系数	54
二、Spearman 秩相关系数	56
三、应用 SPSS 计算 Pearson 和 Spearman 秩相关系数	57
四、应用 PAST 计算 Pearson 相关系数和 Spearman 秩相关系数	58
五、应用 QBASIC 程序计算变量之间的相关性	59
第七章 生态位	61
第一节 生态位概念	61
第二节 生态位指数计算	61
一、生态位宽度	61
二、生态位重叠值	63
第八章 空间格局分析	90
第一节 空间格局的概念	90
第二节 分布格局类型的判定方法	90
一、方差均值比	91
二、 χ^2 检验	91
三、 ϕ 检验	92
四、Morisita 指数	93
五、 C_A 扩散指数	93
六、平均拥挤度	93
第三节 格局分析方法	95
一、单种格局规模分析	95
二、种群分布格局的斑块间隙、斑块大小分析	100
第九章 群落数量分类方法	103
第一节 分类的目的和意义	103

第二节 相似系数和相异系数	103
一、仅适合二元数据的相似关系	103
二、适用于二元数据和数量数据的相似关系	106
第三节 等级聚类方法	108
一、等级聚类的一般步骤	108
二、应用 PCORD 软件进行聚类分析	112
三、应用 PAST 软件进行聚类分析	117
四、应用 SPSS 进行聚类分析	118
第四节 等级划分法	120
一、单元划分法	120
二、双向指示种分析法	122
第五节 非等级分类方法	125
一、相似分类法	125
二、相异分类法	125
三、图论聚类法	126
第十章 排序	129
第一节 排序概述	129
第二节 极点排序	130
一、方法简介	130
二、应用 PCORD 进行极点排序	133
第三节 主成分分析	137
一、主成分分析概述	137
二、主成分分析的原理	137
三、计算步骤	139
四、应用 SPSS 11.0 进行主成分分析	142
五、应用 PAST 2.17 进行主成分分析	144
六、应用 PCORD 进行主成分分析	147
第四节 主坐标排序	149
一、方法介绍	149
二、应用 PAST 进行主坐标排序	150
第五节 对应分析和除趋势对应分析	151
一、对应分析	151
二、除趋势对应分析	154
三、应用 PAST 进行对应分析和除趋势对应分析	156

四、应用 PCORD 进行对应分析和除趋势对应分析	158
第六节 典范对应分析和除趋势典范对应分析	160
一、典范对应分析	160
二、除趋势典范对应分析	165
三、应用 PAST 进行典范对应分析	165
四、应用 Canoco 进行典范对应分析和除趋势典范对应分析	167
五、应用 PCORD 进行典范对应分析和除趋势典范对应分析	172
第七节 无度量多维标定排序	174
一、方法	174
二、应用 PAST 和 PCORD 进行无度量多维标定排序	175
 第十一章 生理生态学中数据统计分析	176
第一节 单因素方差分析	176
第二节 双因素方差分析	180
第三节 独立样本 t 检验和成对样本 t 检验	182
一、独立样本 t 检验	182
二、成对样本 t 检验	183
第四节 重复测量数据的方差分析	184
第五节 实验结果的图表形式	188
一、三线表	188
二、应用 Excel 作图的技巧	189
 第十二章 应用 R 语言进行生态数据分析	193
第一节 R 语言简介	193
一、R 语言的下载和安装	193
二、R 程序的一般运行方法	195
三、R 语言中的基本语句	197
四、R 语言中重要的生态与环境统计学程序包	204
第二节 数据标准化程序	204
第三节 二元数据的相似系数矩阵程序	206
第四节 图解相关系数矩阵的程序	208
一、程序	208
二、运算结果	208
第五节 用 R 语言计算物种多样性指数	211
一、 α 多样性的计算	211

二、 β 多样性计算程序	211
第六节 应用 R 语言进行数量分类	214
一、建立相异系数矩阵的程序	214
二、构建聚类树的程序	214
三、圆形聚类树状图的方法	216
四、聚类和热图的构建方法	218
五、最小生成树	220
第七节 应用 R 语言进行多元回归树构建	223
第八节 C-均值模糊聚类分析	225
第九节 应用 R 语言进行生态位分析	228
第十节 应用 R 语言进行排序分析	229
一、排序程序	229
二、组合聚类 and 排序的方法	234
三、典范对应分析、对应分析、除趋势对应分析和约束排序分析程序	237
四、非度量多维标定	240
五、偏约束排序分析	242
第十一节 R 语言在构建决策树模型中的应用	244
第十二节 应用 R 语言绘制图形的方法介绍	248
一、基本作图语句	248
二、用 R 语言生成热图	251
三、排序二维散点图	252
四、排序的三维散点图	255
参考文献	258
有关软件的参考文献	260

第一章

生态学数据收集

第一节 样方设置

野外生态学调查多数涉及种群、群落和生态系统水平,研究都是从实测数据出发,经过分析、运算,挖掘数据中隐藏的规律,这需要一套标准的数据收集方法。

生态学研究过程中收集数据的过程称为取样(sampling)。实际研究过程中,人们不可能对所研究区域的群落全部进行研究,只能抽取部分进行分析。取样的目的是通过样方的研究去准确地推测群落的总体,抽取的这一部分称为取样单位(sampling unit),包括样方、样圆、样点、样线、样带等,获取样方中生物分布和环境因子数据。生态学取样方法包括主观取样和客观取样两大类。

一、主观取样

主观取样是根据主观判断有意识地选出某些“典型”的、有代表性的样地进行调查。这一取样方法在植被研究实践中曾广泛地使用,它迅速、简便,对有经验的工作者能够取得较好的结果,缺点是不能进行显著性检验。应用主观取样方法获得的数据不能用于统计分析,包括 t 检验、 F 检验、联结、相关或回归等显著性测定,但是适用于如排序等某些多元统计分析。

二、客观取样

1. 随机取样

样方设置是随机的,每一分析对象被抽取的概率是相等的。一般随机取样是将研究地区放入一个垂直坐标中,用成对的随机数作为坐标值,来确定样方的位置。随机数可以取自 Fisher 随机数表(图 1.1)。样方 A 和 B 分别由随机数对(4, 4)和(55, 25)决定。

由随机数决定的样方位置,在实际研究中往往难以确切设置,尤其是在地形复杂、沟

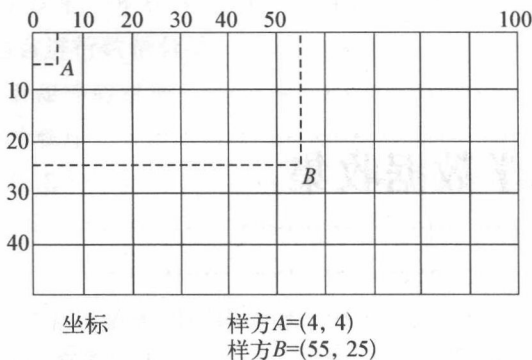


图 1.1 随机取样图示,由随机数确定样方位置(张金屯,1995)

壑交错、裸岩纵横的地方,真正达到随机取样的很少。随机取样的样品可以用于统计分析,从而检验样品的分布是否真正是随机的。

2. 系统取样

系统取样又称规则取样,是根据某一规则系统地设置样方。例如,在山地开展生态调查,先根据野外环境特点,确定第一个样方位置,从低海拔向高海拔,每隔 100 m 设置一个样方。在实际工作中,随机地选择第一个样方后,可以向两个或四个不同方向规则地设置其他样方。取样的方位和间隔的位置需要从环境及群落特点考虑。

3. 限定随机取样

限定随机取样又称系统随机取样,先将研究地段分成大小相等的区组,然后在每一小区组内再随机地设置样方。

4. 分层取样

根据自然界线或不同环境类型,将研究地段分成一些小的地段,再在小地段内进行随机或规则取样,分别代表不同的环境类型。例如,在垂直分布明显的山地,可以将不同植被带作为小地段。分层取样时,小地段的大小并不相等,也难以进行统计分析。

5. 环境因子取样

海拔高度、坡度、坡向、小地形变化等,可以直接从样方中测量记录。土壤等环境因子变化大,还需在样方内再进行随机取样或规则取样。样方中土壤取样时,一般取 5 个点,即样方的中心点和中心点到样方每个角连线的中点。实际测定时,可以将这 5 个样品充分地混合,然后再从中取一部分进行测定,或者将 5 个样品分别测定,对其结果取平均值,后者可以进行统计分析检验。

第二节 样方形状、大小和数目

一、样方的形状

取样的形式有样方、样圆、样点、样线、样带等。为了减少边际效应对测量数据的影

响,样方的边与面积比要小。圆形的周长与面积比最小,但是在野外地形复杂的环境中用圆形进行取样困难很大。方形的边和面积比较小,因而边际影响的误差较小,而且方形容易设置,因此,一般生态学调查中样方为方形。

样线用于一些特殊的研究中,主要用于灌丛和森林群落中,但是使用频率低。格局分析中,由连续样方组成样带是最常用的方法。

二、样方的大小

一般用群落的最小面积作为样方的大小。群落最小面积定义为群落中大多数种类都能出现的最小样方面积,通常用种数面积曲线来确定,即种数面积曲线的转折点所对应的样方面积。由群落最小面积确定的调查样方数据适合多元分析方法,但不适合格局分析。表 1.1 为不同群落类型生态学调查时取样的面积大小。

表 1.1 不同群落类型最小面积经验值

群 落 类 型	群落最小面积/m ²
苔藓群落	0.04~0.25
沙丘草原	1~10
干草原	1~25
草 甸	1~50
高草地	5~50
灌 丛	10~50
温带森林	200~500
热带雨林	500~4000

样方大小在格局分析中非常重要。格局分析中样方大小要明显地小于该类型群落的最小面积。在低矮草地研究中一般用 25~100 cm² 的样方,高草地用 100~625 cm²,灌丛用 0.5~4 m²,森林用 2~25 m²。样方大小选择不当,研究结果均匀分布、随机分布和集群分布就分辨不出来,结果可能它们都是随机分布或均匀分布的。在小格局和微格局分析中,样方会设置得更小。

三、样方的数目

1. 平均数曲线法

当样方数较少时,平均数变化幅度较大,随着样方数目的增加,它的变化幅度逐渐减小,当达到某一样方数目时,它的变化幅度小于允许的范围(如 5% 变化幅度),此时对应的样方数目可以认为是所需要选取的样方数。

2. 面积比法

举例来说,如果研究地总面积为 10 000 m²,样方大小为 5 m×5 m,要求抽取面积为总面积的 5%,即取样面积应为 500 m²,则样方数为 500/25=20。

第三节 无样地取样法

无样地取样一般用于测定森林群落中树种密度、基面积、频度等数据,有最近个体法、最近邻体法、随机配对法和中点四分法(张金屯,1995)。

中点四分法应用得最为广泛,该方法定义为随机样点与每一象限中最近一株间距离的平均值。对于一个样点要测定四个距离(图 1.2),该法要求事先确定好坐标系的方向。

表 1.2 是浙江省遂昌县华西枫杨—枫香群落中 5 个取样点的调查数据,用以说明中点四分法的实际操作方法。通过表中数据,频度 = (含有某种的样点数/样点总数) $\times 100\%$, 以此式计算每一树种在调查地的频度。

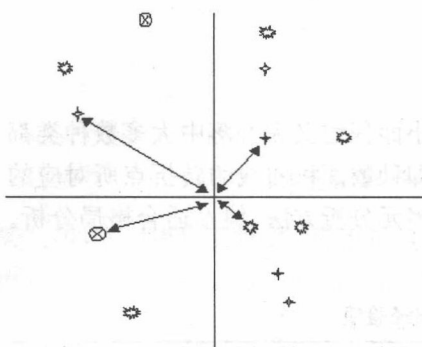


图 1.2 中点四分法调查示意图

表 1.2 中点四分法调查记载表

取 样 点	象 限 号	种 类	距 离/m	基 径/cm
1	1	华西枫杨	0.7	5.5
	2	枫香	1.6	42.5
	3	黄山木兰	3.5	17.0
	4	秀丽槭	2.0	25.0
2	1	华西枫杨	1.1	4.01
	2	华西枫杨	0.8	5.0
	3	华西枫杨	1.9	5.0
	4	华西枫杨	1.8	4.0
3	1	枫香	1.5	4.5
	2	华西枫杨	0.7	3.0
	3	黄山木兰	1.5	9.0
	4	黄山木兰	2.0	23.0
4	1	枫香	3.1	14.0
	2	华西枫杨	1.7	6.0
	3	华西枫杨	1.1	5.0
	4	枫香	1.9	12.0
5	1	枫香	2.5	23.0
	2	枫香	2.2	18.0
	3	华西枫杨	1.4	5.01
	4	黄山木兰	2.8	25.0
合 计	20		35.8	
平均距离/m			35.8/20=1.79	

频度：华西枫杨 $= (5/5) \times 100\% = 100\%$ 、枫香 $= (4/5) \times 100\% = 80\%$ 、黄山木兰 $= (3/5) \times 100\% = 60\%$ 、秀丽槭 $= (1/5) \times 100\% = 20\%$ 。

每公顷株数 $= 10\ 000 / (1.78 \times 1.78) = 3156$ 株，其中，华西枫杨 $= 3155 \times 9/20 = 1420$ 株、枫香 $= 3155 \times 6/20 = 946$ 株、黄山木兰 $= 3155 \times 4/20 = 631$ 株、秀丽槭 $= 3155 \times 1/20 = 158$ 株。

第二章

群落数量和环境特征

第一节 群落的数量特征

一、多度

多度(abundance)是指群落内每种植物的个体数量,对于树木等种类或需要详细研究的群落,常用直接计数法,而对于草本、灌木等个体数目大而体形较小的种类,常用目测计数法。表 2.1 列出了一些具体的估测多度等级的方法。

表 2.1 几种常用的多度等级

Drude 等级	Clements 等级	Braun-Blanquet 等级
Soc. 极多	D 优势	5 非常多
Cop. 3 很多	A 丰盛	4 多
Cop. 2 多	F 常见	3 较多
Cop. 1 尚多	O 偶见	2 较少
Sp. 少	R 稀少	1 少
Sol. 稀少	Vr 很少	+ 很少
Un. 个别		

实测方法可以计测样方中单株个体的数目,也可以根据植物的地上枝条数或丛生植物的丛数等计数,计数时一般以植物的根部是否位于样方内为标准。

二、密度

密度(density)是指单位面积的株数(整个植株或某一部分)。相对密度反映群落内各种植物数目之间的比例,计算方法为

相对密度=某种植物的个体数目 / 全部植物的个体数目 × 100