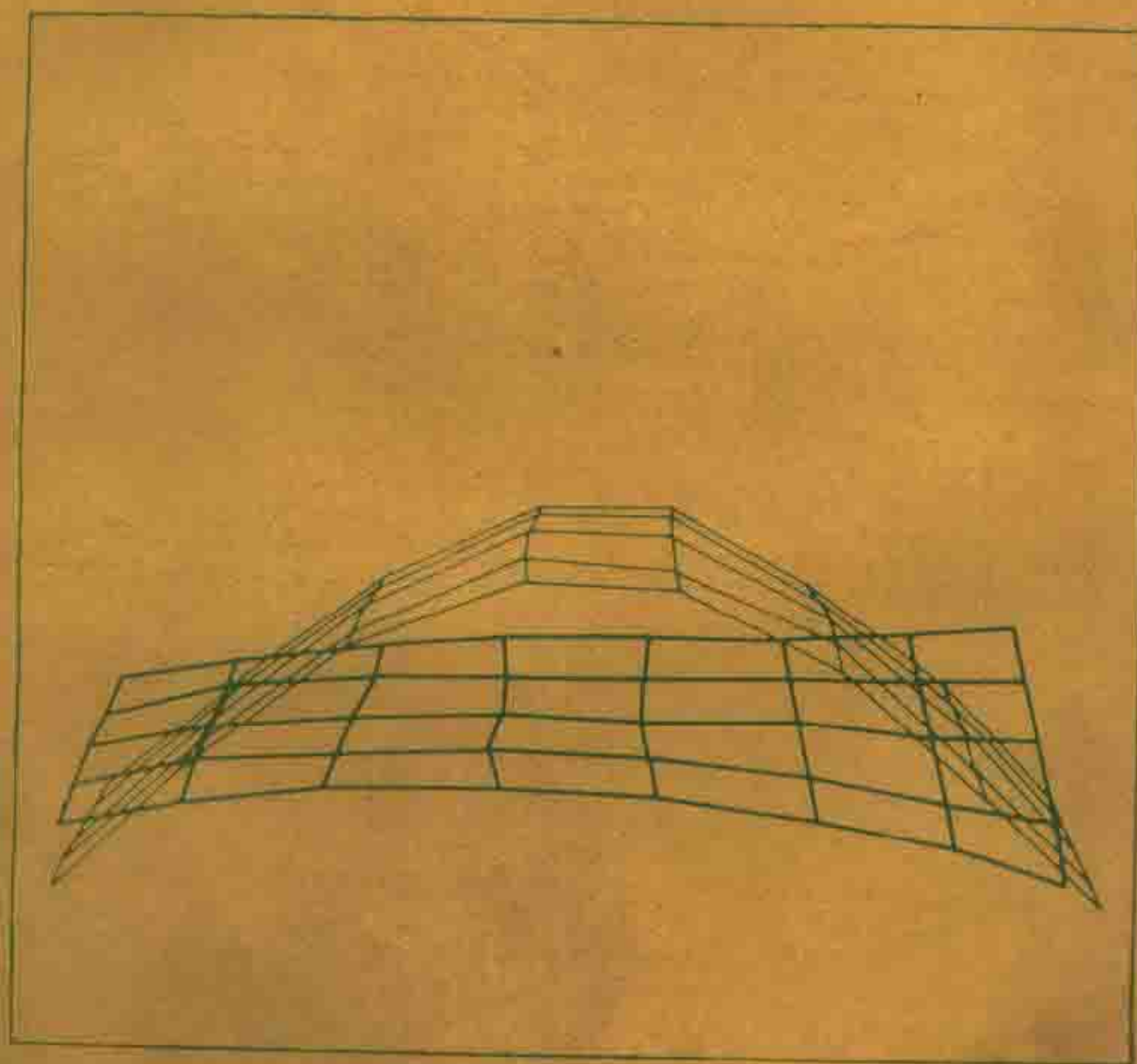


群落生态学中的多元分析

[美] H.G. 小高奇 著



科学出版社

群落生态学中的多元分析

〔美〕 H. G. 小高奇 著

杨 持 杨在中 译
郝敦元 曹泗弟

李 博 郝敦元 校

科 学 出 版 社

1 9 8 9

内 容 简 介

多元分析技术允许归纳大量的、复杂的数据集合,并且提供处理许多由于实际上的约束而不可能实际调查问题的手段。生态学家使用计算机分析植物和动物群落的生态学数据,能够把相似种和相似样方组合在一起,并能概括关于那些影响群落的环境和历史因素的假设。

本书适时地介绍了三种多元分析方法:直接梯度分析、排序和分类。既从理论方面阐述,又从实用观点方面描写;既介绍了传统的方法,又介绍了新的方法,列举了各种各样的例子予以说明。作者提供了该领域的最新综合,这个综合将引起高年级大学生和生态学家们的注意。这些数学工具同样也用于其它领域,从自然资源管理和农学到社会科学和政治学。

本书可供生态学工作者,大专院校有关专业师生以及分类学、古生态学、农学、林学和环境科学方面的工作者参考。

Hugh G. Gauch, Jr.

MULTIVARIATE ANALYSIS IN COMMUNITY ECOLOGY

Cambridge University Press, London, New York, 1982

群落生态学中的多元分析

【美】H. G. 小高奇 著

杨 持 杨在中 译
郝敦元 曾泗第

责任编辑:于 拔

科学出版社出版

北京东黄城根北街16号

中国科学院印刷厂印刷

新华书店北京发行所发行 各地新华书店经售

1989年9月第一版 开本:850×1168 1/32

1989年9月第一次印刷 印张:9 5/8

印数:0001—1 125 字数:251,000

ISBN 7-03-001103-1/Q·171

定 价: 11.70 元

译 者 的 话

近年来,在国外植物生态学研究中已广泛采用多元分析的方法,出版了不少专著。在国内,随着生态学研究的逐步深入,数学方法在生态学研究中的应用愈来愈为人们所认识,工作中也就要求得到更多更新的文献资料及书刊。虽然科学出版社陆续出版了几本译著,出版了阳含熙教授等编著的《植物生态学的数量分类方法》,但由于应用领域的更加广泛,生态学研究队伍的进一步扩大,这方面参考书的需求量也随之增多。

Hugh G. Gauch, Jr. (1982) 编著的这本《群落生态学的多元分析》是 Robert H. Whittaker (1978) 主编的《植物群落排序》、《植物群落分类》著述的补充。该书在内容上相当简炼,在记述方式上,是把直接梯度分析、群落排序和数量分类作为一个整体进行系统的描述,而且既强调了原理又强调了应用,对于初学者来说,易于形成一个完整系统的概念,比起《植物群落排序》、《植物群落分类》两本书来,更易于被初学者接受。

本书共分七章。第一章由杨持译,第二、四章由杨在中译,第三、五章由郝敦元译,第六、七章由曾泗弟译。李博教授与郝敦元校订。

限于水平,在翻译中一定存在不少缺点和错误,欢迎读者批评、指正。

译者

1985.5

为中译本序

多元分析在了解和综合群落生态学中的多变量数据中，已证明是一个有用的工具。我非常感谢内蒙古大学的李博教授为把我的这本书推荐给中国读者所做的工作。我计划在中文版出版以后访问中国，并期望与对多元分析有兴趣的生态学工作者进行讨论、研究。我希望中国的生态学工作者将会发现这些分析是有用的。

H. G. 小高奇

于纽约，伊塔卡，1987.2

序

由于种种原因,综述一下群落生态学中的多元分析是适时的。这个领域近来已得到广泛的发展,有许多方法只是在近几年才提出的。而且,多元分析在应用生态学及其相关领域中正在得到更广泛的使用。最后,这个论题适时的原因还在于急需作出一些生态学问题方法的选择。在解决这些问题时,多元分析是一种主要的研究手段,因为许多生态学问题包含众多的变量、众多的个体或样方(因此包含多元数据),由于实践的限制,很多问题不可能通过实验进行研究。例如,群落数据的多元分析不能代替实验操作,但实验也不能代替多元分析。每个研究方法都有其独到的优点,最强有力的研究策略是使用一套完整的方法。

对本书的编写有两个重要决定。第一,这本书要相当简炼,并且着重多元分析的常用技术。由于比较检验已证明许多被提出的技术的优越性相当少,因此这一策略是合理的。而且,大多数读者主要关心生态学或相关的主题,而不关心方法。这本书是对 Orłóci (1978a) 和 Whittaker (1978a,c) 的较长的且较专门的著述的补充。直接梯度分析、排序和分类作为一套方法提出,并且为了解决研究的问题,把三个方法作为整体进行描述。第二,本书既强调了原理又强调了应用。其基础是与植物和动物群落结构有关的生态学和生物学原理,以及和多元分析的意图和作用有关的数学原理。初步掌握这些原理是进一步掌握所推荐的实际研究方法的基础。

预期的读者包括大专院校高年级大学生、毕业生和生态学专业研究人员,以及同类学科如分类学、古生态学、农学、林学和环境管理等领域的专业研究人员。兼顾到植物、动物和微生物群落;现存的和化石的群落;陆地、水生和海洋群落。现存陆地植物群落给

予特殊的注意,这部分地反映了作者的背景,也部分地反映了这个领域的主要发展。由于多元分析应用极广,因此也简要谈到商业、社会科学和医学等与生态学相隔较远的领域,这些领域的专家们也可以在这里发现其所关心的论题,因为几个多元技术近来在生态学中的发展表现得异常蓬勃,并且异常地节约计算机资源。这本书也将作为一本文献索引。有一个附录记载了有关计算机程序的出处。

我们在康奈尔大学对群落数据多元分析的 10 年研究,主要应归功于已故的 Robert H. Whittaker 的创造性、干劲和洞察力。我衷心地感谢他的指导和鼓励。康奈尔大学生态学和系统分类学教研室给予很大的帮助,犹如学校在计算机和图书资料方面给予的帮助那样。由于 Mark O. Hill 休假访问,对本书的思想贡献极大。John M. Bernard, H. John B. Birks, David J. Hicks, Mark O. Hill, John T. Kunz, Peter L. Marks 和 Robert K. Peet 对手稿提出宝贵的意见; Robert H. Whittaker 审阅了前三章。Steven B. Singer 和 John T. Kunz 提供了有效的技术帮助, Beth H. Marks 和 June Stein 帮助了手稿的准备。为数众多而未能提及的作者对本书所论及的内容作出了贡献,引用他们的工作是对他们应得荣誉的一种表示。我感谢 H. John B. Birks 和剑桥大学出版社约我写这本书。

H. G. 小高奇

纽约, 1981.8

目 录

译者的话

为中译本序

序.....	v
第一章 引论	1
1. 历史回顾.....	2
2. 二维数据矩阵.....	4
3. 群落数据状况.....	7
4. 多元方法.....	12
5. 多元方法的评价.....	34
6. 人为因素.....	36
7. 不断改进.....	41
8. 多元分析的目的.....	43
9. 总的看法.....	46
第二章 取样方法	49
1. 一般考虑.....	50
2. 常规的群落取样程序.....	69
3. 环境和历史资料.....	76
第三章 直接梯度分析	78
1. 基本用途和例子.....	78
2. 方法.....	83
3. 结果.....	84
4. 数据集的性质.....	102
5. 模拟数据.....	108
6. 讨论.....	112
第四章 排序	115

1. 五个概念空间.....	116
2. 排序的目的.....	124
3. 排序技术.....	125
4. 排序技术的评价.....	168
5. 结果的表示和解释.....	172
第五章 分类.....	178
1. 分类的理论.....	179
2. 分类技术.....	186
3. 分类方法的评价.....	207
4. 讨论.....	213
第六章 应用.....	215
1. 一般介绍.....	215
2. 应用群落生态学.....	225
3. 相关的领域.....	236
4. 边远领域.....	241
5. 讨论.....	245
第七章 结论.....	247
附录 通用计算机程序.....	251
参考文献.....	253

第一章 引 论

群落生态学涉及生活在一起的植物和动物的总体，以及和它们发生相互影响的环境和历史因素。群落生态学在生物学领域内作为一个整体最好以主要对象的时空尺度来描述，而不是以诸如概念性见解或方法论等其它标准表述 (Osmond, Björkman 和 Anderson, 1980)。群落生态学家主要以几米到几千米的空间尺度和几周到几百年的时间尺度来研究生命现象。超出这些时空尺度的现象虽然与群落生态学以及它的分支学科的特性有关，但主要由群落生态学家之外的其他生物学家去研究。

群落水平的观点是生态学的主要部分，由于许多理论上和应用上的目的而受到关注。即使一个生态学家的主要兴趣比较特殊时，如研究一个特定种的分布或研究一种特殊污染的影响，他的很多见解也可能来自预先的或并行的群落研究 (Poore, 1956, 1962; Foin 和 Jain, 1977)。

多元分析是处理联立的多个变量的数学分枝：“只要在大量个体上测量的特征多于一个，而且这些特征之间相互作用着，使得有必要对它们进行协同研究时，就需要多变量分析” (Krzanowski, 1972)。由于每个样地是由许多种的多度描述的，还由于许多环境因素影响群落，如此等等，所以群落数据是多元的。多元分析的目的是把多元数据看作一个整体来进行处理，综合这些数据并展示它们的结构。相反，在同一时间处理一个到几个变量的统计学方法对于分析这样的数据，一般是令人烦恼、难以实行和无效的 (Everitt, 1978:5; Williams, 1976:130—6)。多元分析应用于群落生态学是自然的、常规的和富有成效的。

1. 历史回顾

生态学家在世界范围内对植物群落和动物群落进行分析已有几十年了。这些研究是由群落在自然状态下的重要性和吸引力以及作出明智的管理决定的要求所推动。即使注意力明显地偏于环境而不是生物学,但最重要的环境影响也常常施加于动植物群落,因此,群落分析是对环境因素最好的度量和综合。例如,蛇纹岩土壤明显的经济效应,不是直接来自土壤的高镁和低钙水平,而是由于瘠薄和脆弱的植被。

世界性的群落生态学研究导致各种学派的分化,每个学派都有自己的方法论和重点(Whittaker, 1962, 1978a,c; Shimwell, 1971; Maarel, 1979a; Westhoff, 1979)。然而,象所有学科一样,群落生态学的一个共同倾向是朝增加数量方法的方向发展(Cain 和 Castro, 1959:1, 105; Greig-Smith, 1964:210—19; Daubenmire, 1968:39, 268; Orloci, 1978a:2)。这种倾向在过去 20 年间由于生态学家加强了计算机的使用,伴随着多元分析方法的发展而加速了。在数量方法的发展中,无疑有难以使用和不适当的因素,并且很多方法还未经历时间的考验或未证明其现实性,其原因是群落研究中的复杂性和固有的局限性。虽然不能期望原初的研究不经试验和不犯错误就有进展,但比较和评价这些方法确是必要的。群落生态学还是一门年轻的学科,在 1900 年前几乎没有著作(各种群落的偶然的非正式的描述除外),直到 1950 年前后才少量理论的或定量的发展(见 Greig-Smith, 1980 年的综述)。当前的成就是显著的,主要考虑的是近 30 年的工作(Goodall, 1962; Egerton, 1976; McIntosh, 1976, 1980)。

对群落生态学作定量研究有几个优点:(1)数量方法“要求使用者在问题的表述,预见所收集的数据类型的含意,所使用的取样计划和所选择的数据分析以及统计推断方法诸方面都要准确而明晰”(Orl6ci, 1978a:2; 还看 Greig-Smith, 1964:212)。(2)

从使用者方面来讲,数量方法“要求在数据收集、分析和推理的全过程中执行操作规程和作出决定的均匀性与一致性”(Orl6ci,1978a:2)。这种一致性有利于生态学家之间以及与其它领域的科学家间的合作和交流(Orl6ci,1978a:3)。(3)“定量方法可以测定和评价较小的差别”(Greig-Smith,1964:212)。(4)由于定量方法的均匀性和一致性,格式化的方法“自然适合于计算机处理”,这种处理提供了某些逻辑上的便利(Orl6ci,1978a:2)。

由于这些理由,数量方法能用来解决非格式化的方法简直不能回答的问题(Greig-Smith,1964:212; Williams,1976)。尽管有这些优点,数量方法并不一定需要比非格式化方法付出更大的努力。事实上,“以一种特定的方法只做一件事的习惯可能是有效的”(Cain 和 Castro,1959:105),并且计算机处理能节约很多花费在无味的和易造成误差的工作上的时间。当一项研究数据的复杂性或数据总量增加时,这些优点更易被觉察。

仅在几年前,对生态学家有用的多元方法还不如现在那样强而有力。曾存在三个问题:(1)真实地比较日益增多的多元方法,需要对同一数据集合应用多种方法,但这是缺乏的。尽管存在较好的方法,但常常由于选择失当而导致各种方法评价的相互抵触的论断,从而纠缠不清。(2)群落的数学性质需要明晰而准确地被模拟,并把所得到的群落模型同多元方法的基础模型比较,以便认识其不足之处,从而发展更适宜的方法(Noy-Meir 和 Austin,1970; Swan,1970; Austin 和 Noy-Meir,1971; Gauch 和 Whittaker,1972a, b, 1976; Gauch, Whittaker 和 Wentworth,1977; Austin,1980b; Hill 和 Gauch,1980)。数据分析本身的目标亦须明确,以便促进向这个目标前进(Dale,1975)。(3)大多数使用计算机的多元方法要求按平方或立方(或更多)数量级而增加数据(至少在执行有效的计算程序时如此)。因此,许多方法,甚至用计算机的方法也仅限于处理数百或乃至数十个样本。

现在,有效的、高效的和适用的多元方法对群落数据的排序、

等级分类和非等级分类都是可用的。较新的方法也很可靠,因此,现在推荐使用多元方法比起前几年来就爽快多了。群落数据的内在限制因素和计算效率限度的理论评价指出,这些方法中的某一些可以接近理论限度,因而性能和速率都不会再显著地改善。所以,现在来评述群落生态学中的多元方法是适时的。

2. 二维数据矩阵

群落研究涉及到从大量样方里观测许多个种的多度。所得的数据可以排列在一个称为群落表的矩阵之中,即一个原始数据矩阵或简言之,一个数据矩阵(如关于多元方法一节里的表 1.1)。数据矩阵有二维结构,种在一边,样本在另一边。通常在二维的样本-种的数据矩阵里,行表示种,列表示样本。

样本-种矩阵是个体-属性矩阵大类中一类特定的矩阵。事实上,科技人员最经常关心的是属性多于 1 的情况,因而多数数据是多元的。而且,由于经常关心的是 1 个以上的个体,具二维的个体-属性矩阵格式的数据比其它任何格式就更具有科学意义。多元方法在全世界爆发性的发展是这种资料普遍性的一种反应。在第六章里将全面回顾多元分析的来龙去脉,而下面列出的清单指出了生态学和其它学科领域里二维数据的某种普遍性。

生态学

 样本-种

 生境-环境参数

 种-生态位 (niches)

 种-行为格局

分类学

 标本-特征

遗传学

 变种-遗传特征

农学

 土壤-特征

古生物学和地质学

样本-种

样本-地球化学参数

考古学

生境-表象

语言学

人-词汇

语言-语音

心理学

人-特性

民族-性格

商业

人-消费选择

公司-产品

因为多元方法的应用各式各样,所以著作多而分布广泛。用于生态学的许多方法是由分类学、心理学、考古学和商业工作者开创或发展的,因为这些不同领域来的数据有许多共同之处,因而方法的交流被证明是有益的 (Simon, 1962; Crovello 1970; Bflashfield 和 Aldenderfer, 1978)。然而,每个领域又有它特殊的数据性质和分析要求,以致每个领域都需要检验不同方法对其自身目的的适宜性。在每个领域里最普遍使用的多元方法大大受到前面介绍的特殊方法的影响。例如,流行于生态学和分类学的主分量分析,以及心理学和商业里的无度量多维计数。遗憾的是,许多选择仅反映习惯而不是性能,但是,对于一个给定的领域,要去判明这许多方法的有效性则是花费时间的。

有时观察数据集合最自然的方式是三维(或多维)矩阵,而不是通常的二维矩阵。例如,在具季节性的海洋底栖生物的研究中,Williams 和 Stephenson (1973) 有一个样本-种-时间数据集合。类似的,由山区草地的永久样区,Swaine 和 Greig-Smith (1980) 分析了一个生境-种-时间-处理-重复的数据集合。然而,排序和分类的计算只能接受二维数据矩阵作为输入[某些三维无度量多

维计数模型除外,这种多维模型在多数生态学应用中可能是无效的(Young 和 Lewycky, 1979:11, 13)[]。如果要利用排序和分类,多维数据必须整理,归并,或划分为一个或几个二维矩阵(Crovello, 1970; Williams 和 Stephenson, 1973; Gillard, 1976; Williams 和 Edye, 1976; Jöreskog, Klován 和 Reymont 1976: 117—18; Austin, 1977; Green, 1979: 69; Maarel, 1980b; Nishisato, 1980:149—63; Swaine 和 Greig-Smith, 1980)。很难说,重要方法上的改进会能更好地处理多维数据,还是这种内在复杂的数据,而人们又要求使用不多于二维或三维的变量来描述,于是便提出了一些不可避免的问题,这些问题只好尽其可能用刚刚列出的文献中所使用的方法去应付。在任何情况下,随着一个数据集合复杂性的增加,要求规定一个稳定增加的数据量,即使分析技术是有效的,也还需要大量的数据资料。

人们认为足以判明一个多元方法的个体或属性的数量阈值是任意的,并且某种程度上随数据集合的性质而变;然而,可以大致取作 10 到 15。较少的变量,特别是五个或更少,对多元方法是不适宜的,而对方差分析和分散区组的统计分析可能更富有成效。可是,一个例外情形关系到多元分析对某些矩阵的有效性,例如: 1000×5 , 这种矩阵虽然有一维很短,但另一维是很长的。

这本书里介绍的多元分析适用于这样的数据集合: (1) 构成二维矩阵; (2) 矩阵至少在 10×10 或 15×15 左右。这两个要求必须满足。在这本书中论述的主要实例是样本-种群落数据矩阵,但各种类型二维数据的结构和分析有许多共同的特征,使得多数多元方法广泛关联各种各样的问题。当然,当数据集合性质和采取的多元技术相匹配时,才可望得到有用的结果,至少部分是这样。如果数据值基本上是随机数,或者如果它们的结构与技术模型很不相同,多元分析将是无效的。数据性质和分析模型之间的合理匹配,可以被认为是获得有效结果的第三个要求。前两个要求极易判定;第三个要求在某些情况下是不重要的,但在另一些情况下,最简单的方法可以考验一个多元分析,并且看出其结果是

否有用。

3. 群落数据状况

一旦建立起数据矩阵,分析一个样本-种数据矩阵涉及数据的四个方面:随机干扰、重复性、关系和分离样方(outliers)。

(1) 随机干扰

重复的群落样方(从同一小块同质的群落里取的不同样方)是难得恒等的。确切地说,重复样方间平均有 50—90% 的相似性,这种相似性依赖于样方大小,测量或估计的精密程度,种的数目等等(Dahl 1960; Maarel, Janssen 和 Louppen, 1978; Gauch, 1981)。一个数据集中的样方在种类组成方面的变化,部分地归因于环境因素中令人关注的变化,附带地也归因于不受人注意的随机干扰。数据“受随机干扰”的意思在于,来自相同环境条件的样方在种类组成方面不尽相同,因而,数据部分地反映令人感兴趣的结构,部分地反映随机干扰。

随机干扰的生物学原因是复杂的,包括个体偶然的分配和组合,动物的活动,局部的扰动,以及样方面积尺度以下的环境异质性。干扰也产生于有限个样方的统计局限性和测定或估计种多度的局限性。

在群落的多元分析研究中,随机干扰的有效定义是,一个种的多度变化同另一个种的多度变化不协同(Poore, 1956; Gauch, 1981; 也看 Green, 1977)。用“协同”而不用相关,是意味着以一种由多元分析所揭示的系统方式的关联,而不一定是一种简单的会给出较大的相关系数的线性方式的关联。相反,数据结构包含同时发生的许多种的协同改变(Pignatti, 1980)。协同暗示了许多种以一种有意义的和可能解释的方式对一种(或其它种)环境因素的反应;干扰则意味着不协同和种的多度方面显然是偶然出现的差别。协同性可用一个连续的等级谱描写,因此上面的定义可以精确化

为：干扰是一个种的多度变化与另一些种的多度变化间的协同性显著地小于那些已被观测到的较大的协同性。划定较大的协同性的界限牵涉一个主观决定，野外经验会使这种决定作得更准确。

干扰的这个定义是不精确的，就其干扰的本性是不顺从精确的处理。然而对于常规目的，即对干扰的兴趣较低时，这个定义是适当的。应该看到，干扰的概念产生于数据和分析间的相互作用。除某些特殊分析和研究目的外，定义并没有什么意义，因此也就没有结构和干扰之间的区别 (Simberloff, 1980; Wimsatt, 1980; 还看 Webb, Lasleski 和 Bernabo, 1978)。

数据集合的全部信息可以认为是 $X\%$ 的干扰和 $(100-X)\%$ 的结构，这里的 X 通常为 $10-15\%$ [以位 (bits)、方差或无论那种单位来表示信息总量]。多元分析的目标是要尽可能忠实地表达数据结构，而最小地表达干扰。因此，理想的是概括原始信息的 $(100-X)\%$ ，而且有选择地反映消除了干扰的数据结构信息 (Greig-Smith 1971; Prentice, 1980a)。如果 100% 地恢复原始信息，在减少干扰方面，分析将完全失败。

为了实用上的原因，排序结果往往被限制在 1 到 3 维的范围，而分类结果被限制在 10 到 50 个聚类，因此要恢复全部感兴趣的数据结构多半是不可能的。这种情况下，在结果形式的选择上，目标是要尽可能多的表达重要的协同数据结构，而尽可能不去管随机干扰。

(2) 重复性

一个样本-种数据矩阵通常有适当的甚至较大程度的重复：许多样方在种的组成上与另一些样方非常相象，并且许多种在样方中出现的情况与另一些种也很相象 (Moore, Fitzsimons, Lambe, 和 White, 1970; Orłóci, 1974a, 1975, 1978a:17-19; Gauch, 1980; 还看 Simon, 1962)。重复包含了协同的种反应和相似样方；而随机干扰则相反。