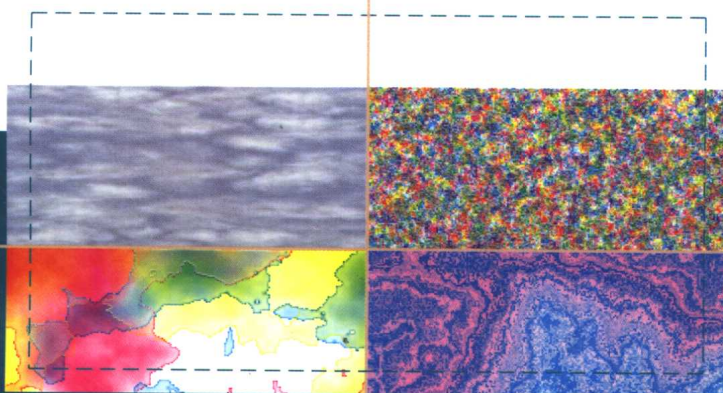


地统计学及在生态学中的应用

王政权 编著



科学出版社

内 容 简 介

本书系统地介绍了地统计学的基本理论、方法及在生态学中的应用。本书共分7章。除第1章绪论外,第2章和第3章介绍了抽样数据的统计分析方法及空间数据的表示。第4章、第5章和第6章介绍了地统计学中区域化变量理论、变异函数的结构分析和空间局部估计的基本理论和方法。第7章介绍了地统计学在生态学中的应用实例。在给出基本概念和公式的同时,附有计算实例,使读者更好地掌握地统计学的基本理论和方法,在应用过程中不致于产生困难。每章给出有关参考文献,供读者进一步阅读。

本书要求读者具备线性代数、数理统计的基本知识,可供植物生态学、景观生态学、生物学、地学、土壤学和农林院校方面的科技人员,有关专业的大学生和研究参考。

图书在版编目(CIP)数据

地统计学及在生态学中的应用/王政权编著. -北京:
科学出版社, 1999. 2

ISBN 7-03-006833-5

I. 地… II. 王… III. 生态环境-统计学 IV. Q141

中国版本图书馆 CIP 数据核字(98)第 17729 号

科 学 出 版 社 出 版

北京东黄城根北街16号

邮政编码:100717

科 地 亚 印 刷 厂 印 刷

新华书店北京发行所发行 各地新华书店经售

*

1999年2月第 一 版 开本: 787×1092 1/16

1999年2月第一次印刷 印张: 12 3/4

印数: 1—15 000

字数: 292 000

定价: 26.00 元

(如有印装质量问题,我社负责调换〈新欣〉)

序

生态学正以前所未有的速度向深度和广度发展,扩大研究范围,加深理论认识。景观生态学的崛起和快速发展,正反应了生态学的这一总的趋势。

生态学空间问题的研究,是生态学理论研究的重要问题。在不同尺度上研究空间格局对生态学过程的影响,并预测其变化,不仅是景观生态学,而且是整个生态学研究的核心任务。空间异质性是导致格局形成的主要原因,空间异质性又随尺度而改变。当研究尺度发生变化时,格局与过程也因此而不同。以往的生态学研究忽视了空间异质性与尺度对格局和作用的作用,使它不能有效地揭示生态学规律(至少在较大尺度上如此),并为解决某些现实问题服务。要深刻分析和准确表达空间异质性,需要一种数学工具,地统计学(geostatistics)正满足了这一需要。地统计学提供了一种有效地分析和解释空间数据的方法。生态学家利用这一方法可分析、认识和解释与空间异质性有关的复杂现象,建立空间预测模型,并进行空间数据插值和估计,达到对生态学空间研究的目的。

地统计学诞生于60年代。我国于70年代末首先在地质采矿界加以介绍。90年代以来,国外许多生态学家采用地统计学方法研究生态学中的空间异质性与格局等问题。在国内生态学界阐述地统计学的价值,并加以系统介绍的应推王政权博士的这本《地统计学及在生态学中的应用》。书中介绍了地统计学的基本概念、方法和理论。从样本数据的统计分析、空间分析、区域化变量理论、变异函数的结构分析,直到空间局部估计的克立格法等等。本书取材精当,叙述简明,不仅景观生态学工作者应当细读,其他生态学分支的工作者无疑也会从中得到启发并学到很多东西。

书的后半部分,作者详细介绍了阿拉斯加北极冻原,黑龙江省伊春市带岭凉水自然保护区的阔叶红松林及森林土壤养分,水生甲壳类动物种群,黑龙江帽儿山实验林场的落叶松毛虫等6个案例。其中凉水和帽儿山两处是作者亲自野外工作的成果,具有示范性。作者通过在这两处的研究,还得出一些具有潜在价值的规律。例如,在小尺度上,红松分布没有明显的规律性,而在中尺度上具有群团状分布的特点。凉生的阔叶红松林具有高度的空间异质性,红松的空间异质性与其它树种有较大的差异,云杉、冷杉和枫桦的标准化变异函数曲线相似。但阔叶红松林中的每一树种在不同尺度上的空间格局具有相似性。帽儿山实验林场的落叶松毛虫发生初期、中期和末期空间异质性不同。中期较高,而且主要由空间自相关因素引起,初期和末期较低。落叶松毛虫种群的空间格局在发生的初期、中期和末期具有相似性,不论初期、中期和末期,种群聚集的分布型不变。在落叶松毛虫发生初期,种群斑块较小,而发生的中期和末期,种群斑块较大。表明随着时间增加,种群数量和危害面积也在增大,等等。这些情况的阐明,无疑会对实践有所帮助。

可以预料,本书的出版和发行,将会引起我国生态学界和地学界的很大兴趣,并对推动景观生态学在我国进一步的发展发挥作用。

陈昌笃

1998年7月于北京

目 录

序

第1章 绪论	(1)
§ 1.1 地统计学的概念及发展历史	(1)
§ 1.2 地统计学与经典统计学的区别及在生态学中的应用	(2)
§ 1.3 本书的目的及内容	(4)
参考文献	(5)
第2章 样本数据的统计分析	(8)
§ 2.1 总体、样本和随机样本	(8)
§ 2.2 频数分布	(9)
§ 2.3 统计特征数	(13)
§ 2.4 两个样本之间的统计	(18)
参考文献	(25)
第3章 样本数据的空间分析	(26)
§ 3.1 空间数据的图形表示	(26)
§ 3.2 空间数据的相关统计	(29)
参考文献	(34)
第4章 区域化变量理论	(35)
§ 4.1 区域化变量的概念及性质	(35)
§ 4.2 协方差函数和变异函数	(38)
§ 4.3 协方差函数和变异函数的计算公式	(43)
§ 4.4 方差估计理论	(52)
§ 4.5 正则化变量及性质	(62)
参考文献	(64)
第5章 变异函数的结构分析	(65)
§ 5.1 协方差函数和变异函数的性质	(65)
§ 5.2 变异函数的理论模型	(69)
§ 5.3 变异函数的结构分析	(74)
§ 5.4 变异函数理论模型的最优拟合	(88)
§ 5.5 结构分析的基本步骤	(98)
参考文献	(101)
第6章 空间局部估计	(102)
§ 6.1 克立格法概述	(102)
§ 6.2 普通克立格法	(104)
§ 6.3 协同克立格法	(133)

§ 6.4 泛克立格法	(142)
参考文献.....	(149)
第 7 章 地统计学在生态学中的应用.....	(150)
§ 7.1 空间异质性定量分析方法	(150)
§ 7.2 阿拉斯加北极冻原景观空间异质性研究	(156)
§ 7.3 阔叶红松林空间异质性	(161)
§ 7.4 空间分布格局定量分析	(168)
§ 7.5 森林土壤养分的空间异质性与格局	(177)
§ 7.6 水生甲壳动物种群的空间变异与格局	(185)
§ 7.7 落叶松毛虫种群空间格局的变化	(189)
参考文献.....	(193)
结束语.....	(196)

第1章 绪论

- 地统计学的概念及发展历史
- 地统计学与经典统计学的区别及在生态学中应用
- 本书的目的及内容

§ 1.1 地统计学的概念及发展历史

§ 1.1.1 地统计学的概念

地统计学(Geostatistics),亦称地质统计学,于50年代初开始形成,60年代,在法国著名统计学家 G. Matheron 的大量理论研究工作基础上形成一门新的统计学分支。由于它首先是在地学领域,如采矿学、地质学等中发展和应用的,因此得名地统计学。

“地统计学”一词首先是 G. Matheron(1962)采用的,他在1962年给地统计学下过一个较早的定义:“地统计学即以随机函数的形式体系在勘查与估计自然现象中的应用。”随着地统计学的发展,到了1970年,他又将地统计学定义为:“是以区域化变量理论在评估矿床上的应用(包括采用的各种方法和技术)。”但是地统计学近20年的发展表明,它不仅仅在地质学中应用,而且在土壤、农业、气象、海洋、生态、森林和环境治理方面也开始应用,因此,一些地统计学工作者(Webster 1985,王仁铎等 1987,Issaks 等 1989,侯景儒等 1993)认为:“地统计学是以区域化变量理论为基础,以变异函数为主要工具,研究那些在空间分布上既有随机性又有结构性,或空间相关和依赖性的自然现象的科学。”因此,凡是要研究空间分布数据的结构性和随机性,或空间相关性和依赖性,或空间格局与变异,并对这些数据进行最优无偏内插估计,或要模拟这些数据的离散性、波动性时,均可应用地统计学的理论及相应的方法(侯景儒等 1993)。

§ 1.1.2 地统计学发展简史

统计学在地质科学中的应用在一百多年以前就开始了,主要是根据地质勘探所取得的大量信息,对地质矿产资源进行定量估计和评价。但是在40年代末期,一些地质学家和为地质矿产资源进行定量估计和评价的统计学家们发现经典的统计学由于没有考虑到地质样品或块段体积的大小,没有考虑到品位的空间变化特征,没有考虑到品位之间的空间相关性及估计精度的概念,因而在地质估计和评价存在许多不能克服的缺陷(王仁铎等 1987)。后来发现,地质变量不是纯随机变量,因而不能用简单的统计方法去估计和评价。在此同时,数理统计的研究出现了自相关和变异函数的概念,找到了空间或时间连续或相关的变量的数据处理的基本方法。在40年代末和50年代初,南非的矿山工程师 D. G.

Krige 和 H. S. Sichel 从南非金矿储量计算的具体问题出发,提出了用样品的空间位置和相关程度来估计块段品位及储量,使其估计误差最小的储量计算方法,即 Kriging 的初步(侯景儒等 1993)。随后,法国的在地质矿产资源估计方面优秀的统计学家 G. Matheron,在 Krige 和 Sichel 研究基础上进行了大量理论和实际研究,于 1962 年提出了区域化变量概念,并出版了《应用地统计学导论》的专著。紧接着于 1963 年在《经济地质》杂志上发表了“地统计学原理”一文,将变异函数和空间局部估计的内在联系通过区域化变量理论贯穿起来,系统地论述了地统计学的基本理论、方法及应用,为地统计学奠定了理论基础。从此以后,地统计学的理论和方法才在地质学领域引起足够重视。G. Matheron 的最大贡献是将早期零散的研究归纳、整理和系统化。因此,大多数地统计学研究者都一致认为地统计学的主要理论是由 G. Matheron 创立和发展的(David 1977, Journel 等 1978, Webster 1985, Issaks 等 1989, Cressie 1991)。

在 60 年代和 70 年代,地统计学的研究主要是进一步完善和改进理论和方法。在这期间,各种空间局部估计的方法,如普通克立格法,泛克立格法,折取克立格法等初步形成并在地质学领域得到广泛应用(侯景儒等 1993)。在 70 年代末及 80 年代初,一批地统计学研究理论和应用的专著相继出版。如 M. David 的《矿产储量地统计学评价》(1977), A. G. Journel 等的《采矿地统计学》(1978), I. Clark 的《实用地统计学》(1979), B. D. Ripley 的《空间统计学》等。其中 Journel 等的《采矿地统计学》一书影响较大,成为 80 年代和 90 年代地统计学理论和应用研究的经典参考书。该书全面总结了 80 年代以前的地统计学理论研究和应用研究的成果。到 80 年代末 90 年代初,在美国又出版了两本地统计学专著,一本是 E. H. Issaks 和 R. M. Srivostava 合著的《应用地统计学导论》(1989),另一本是 N. Cressie 的《空间数据统计学》(1991)。前一本书主要介绍地统计学应用的方法,通俗易懂。后一本书偏重理论证明和推导,并附有大量实例。

在我国 1977 年开始介绍地统计学。1982 年,侯景儒等首先将 A. G. Journel 等人的专著《采矿地统计学》(中文名称为《矿业地质统计学》)译成中文,1987 年王仁铎等出版了《线性地质统计学》一书作为高等学校教材。1989 年孙惠文等翻译出版了 M. David 的《矿产储量地质统计学评价》。1993 年,侯景儒等出版了《矿床统计预测及地质统计学的理论和应用》一书。上述几本专著和译著在中国的出版,为地统计学在中国的理论研究和应用研究打下了基础。

§ 1.2 地统计学与经典统计学的区别及在生态学中的应用

§ 1.2.1 地统计学与经典统计学的区别

数理统计是以概率论为基础的一门研究随机现象统计规律的应用数学学科。地统计学与经典统计学存在较大区别:①经典统计学研究的变量必须是纯随机变量。该随机变量的取值按某种概率分布而变化。而地统计学研究的变量不是纯随机变量,而是区域化变量。该区域化变量根据其在某一域内的空间位置取不同的值,它是随机变量与位置有关的随机函数。因此,地统计学中的区域化变量既有随机性又有结构性。②经典统计学所研究的变量理论上可无限次重复或进行大量重复观测试验。而地统计学研究的变量则不能进

行这样的重复试验。因为区域化变量一旦在某一空间位置上取得一样品后,就不可能在同一位置再次取到该样品,即区域化变量取值仅有一次。③经典统计学的每次抽样必须独立进行,要求样本中各个取值之间相互独立。而地统计学中的区域化变量是在空间不同位置取样,因而,两个相邻样品中的值不一定保持独立,具有某种程度的空间相关性。④经典统计学以频率分布图为基础研究样本的各种数字特征。地统计学除了要考虑样本的数字特征外,更主要的是研究区域化变量的空间分布特征。因此,地统计学的主要研究是围绕着变量空间分布理论和估计方法。正是上述主要区别,导致地统计学研究与经典统计学相比具有较多的优点和特色,并在具体实践中迅速发展。

§ 1.2.2 地统计学在生态学中应用

在地学领域,地统计学数据主要来自所研究的区域上的抽样,分析各种自然现象的空间变异规律和空间格局,并已被证明是研究空间变异和空间格局的有效方法(李哈滨等 1992)。然而在 70 年代以前,地统计学的应用基本局限在地质学领域。从 70 年代以后,地统计学被应用于土壤学和水资源研究。J. B. Campbell(1978)在研究两个土壤制图单元中砂粒含量和 pH 空间变异时,首先采用地统计学。随后, Burgess 等(1980a, 1980b, 1981, 1984), Burrough(1983a, 1983b), Webster(1985), Tranggmar 等(1985)将地统计学引入土壤科学研究,并做了大量介绍和实例研究。到目前这方面研究已成为土壤科学研究的重要内容之一(沈思渊 1989),并取得许多重要的科研成果。

格局与过程始终是生态学研究的核心。种群和群落的空间变异的问题很早就被生态学家注意到(Watt 1947, Goodall 1952, Whittaker 1956, Greig-Smith 1957, Pielou 1957, 1959)。当时主要是根据经典统计学或多元统计分析的方法研究种群和群落的数量分类和排序问题,同时也用于种群和群落的格局分析。1964 年, Greig-Smith 出版了他的名著《定量植物生态学》(第二版),重点地描述了格局分析方法,并对已有的数量分类和排序方法进行了总结。在 70 年代,随着电子计算机的普及,生态学家们(Whittaker 1973, Kershaw 1973, Orloci 1978)对种群和群落的数量分类、排序、格局进行了大量实际研究,形成较完整的种群和群落数量分类与排序生态学分支。但是生态学家们并不因此而满足,在研究中他们提出“现实的格局是怎样形成的?它将来的发展会是什么样?为什么同一种方法只要调查的群落面积大小不一样,得到的空间格局会不同?”“为什么在 60 年代和 70 年代种群生态学、群落生态学、生态系统生态学研究形成的理论和方法,在解决生态学实际问题时发生了困难?”一些理论及方法也受到前所未有的质疑和挑战(McIntosh 1987, Harbert 1984)。面对这些问题,北美的一些优秀生态学家们(Tilman 1984, 1989, Franklin 1987, Sokala 等 1987)进行了认真的思考。主要原因是生态格局与生态学过程有关,生态格局随空间尺度变化,空间异质性是导致格局的主要原因。而以往的研究忽视了空间异质性与尺度对格局与过程的作用。随着生态学理论研究的深入,尤其是景观生态学的发展,许多生态学家认识到空间异质性在许多生态学理论中起中心作用(Legendre 等 1989, Turner 1991, Sokala 1991, Levin 1992, Kareiva 1994),并引起高度重视(Dutilleul 等 1993),成为 90 年代生态学理论研究的新的范式(Li 等 1994)。

空间异质性是指系统或系统属性在空间上的复杂性和变异程度,包括系统属性的空间组成,空间构型和空间相关。它在生物学系统各个层次上都存在,是许多基本生态学过

程和物理过程在时间和空间连续系统上长期作用的结果(李哈滨等 1992)。鉴于地统计学的特点及在土壤空间异质性和格局研究中的优越性, Philips 于 1985 年首先采用地统计学方法对美国新泽西州的芦苇群落复杂性进行了研究。随后, Robertson(1987)在“Ecology”上介绍了地统计学在生态学研究中的两个实例。同时, Burrough(1987)在 Jongman 等主编的《群落和景观生态学中的数据分析》专著中系统地介绍了变异函数分析方法和空间局部估计方法在生态学中的应用。Kemp 等(1989), Schotzko(1989)采用地统计学方法分别研究了昆虫种群的空间变异和格局。1989 年, Legendre 等在“Vegetatio”发表了《空间格局和生态学分析》长文, 充分肯定了地统计学在空间异质性与格局分析中的重要作用。紧接着, Turner(1991)将地统计学方法应用于景观生态学的定量研究。Rossi 等(1992)和 Dutilleul 等(1993)分别在“Ecological Monographs”和“Oikos”发表了地统计学在生态学应用方面的两篇经典文献。作者用大量的篇幅和实例, 系统地介绍了地统计学在动物、植物、微生物种群或群落空间异质性, 空间格局与过程研究中的基本理论, 基本方法及应用。并阐述了地统计学方法与过去方法的主要差别、优越性和特点。1993 年在美国 78 届生态学年会上, 美国生态学会以“地统计学与生态学”为主题, 专门邀请 11 位生态学家作大会发言, 阐明地统计学在生态学理论和实际研究中的潜力和前景。

生态学是研究生命有机体与环境间关系的科学。生命有机体常常以个体、种群和群落的形式, 分布在某种特定的空间上或时间上, 它们具有高度的空间和时间异质性。经典统计学由于其基本假设的限制, 在研究个体、种群和群落空间异质性或空间自相关方面具有较多的缺陷。地统计学为生态学家提供了一个非常有效的分析和解释空间数据的方法 (Robertson 1987, Rossi 等 1992)。具体地, 它可以为生态学家在研究生态学各种复杂问题时提供: ①定量地描述和解释空间异质性或空间相关的方法; ②建立各种有关的空间预测模型, 并进行空间数据插值和估计; ③对空间格局的尺度、几何形状、变异方向进行定量地分析和有效地估计, 并将空间格局与生态学过程联系起来; ④它为生态学家在各种尺度上进行空间抽样时, 提供最优的抽样方法; ⑤它可以帮助景观生态学家建立景观模型, 并进行景观模拟; ⑥环境因子的地统计学分析有助于生态学家更深刻地了解生命有机体(个体、种群和群落)空间变异的机制。最近几年的研究表明, 地统计学无论是在土壤的空间格局 (Robertson 等 1993, 1995, 1997, Schlesinger 等 1996), 植物群落的空间变异 (Jackson 等 1993, Katherine 等 1995), 还是在景观 (Li 等 1995, Leadley 1996) 干扰格局等方面的研究, 都显示出经典统计学不可替代的优越性。

§ 1.3 本书的目的及内容

§ 1.3.1 本书的目的

P. Kareiva(1994)指出, “空间问题是目前生态学理论研究的前沿”。地统计理论和应用已被扩展到分析各种自然现象的空间格局, 已被证明是研究空间变异的有效方法。应用地统计学研究生态学有关空间问题, 已成为生态学理论研究的主要内容 (Halvorson 1994)。目前在国内外出版的地统计学的几本专著, 基本上都是在地质学领域内的, 在生态学领域里的应用主要是一些研究论文, 而国内应用地统计学进行有关生态学研究的论

文更少。因此,本书的主要目的是向生态学工作者系统地介绍地统计学的基本理论,基本方法及在生态中的应用实例,满足生态学工作者在研究有关生态学空间问题,尤其是空间异质性、格局、尺度等问题时的需要。

应该看到,地统计学还是处于发展之中的一门应用统计学的一个分支,许多新的理论和方法不断地被地统计学家提出,能否应用于生态学的研究要靠生态学家自己去检验。本书所介绍的基本理论和方法仅是一些在生态学应用中比较成熟的内容。读者通过本书,可了解地统计学的框架,达到“抛砖引玉”的作用。这里还要提请生态学工作者注意的是,地统计学也并不是万能的,它只是一个研究自然空间现象问题时的一个工具,许多生态学中有关空间问题的最终解决要靠生态学本身的研究。

§ 1.3.2 本书的内容

全书共分7章。除第1章绪论外,第2章主要介绍抽样数据的统计分析。包括单一变量和二个变量之间的数据统计分析。这一章的内容与经典统计学的内容一致。第3章主要介绍具有空间位置的抽样数据空间表示方法。包括数据的空间展点,等直线图,符号图,指示图及空间相关的统计。使读者对空间数据的分析有一个初步印象。第4章主要介绍区域化变量理论。包括相关函数,协方差函数,变异函数,估计方差和离散方差等。第5章主要介绍变异函数的结构分析。包括变异函数的性质、功能,各种变异函数的理论模型,变异函数的最优拟合及结构分析的基本步骤。这两章强调了变异函数的性质和进行结构分析时采用的技术。第6章主要介绍空间局部估计的方法。包括普通克立格法(点克立格法和块段克立格法),协同克立格法及泛克立格法。在普通克立格法中,重点介绍了影响普通克立格法估计精度的各种因素,及在空间局部估计中要注意的一些问题。第7章主要介绍地统计学在生态学中的应用。包括空间异质性定量研究的变异函数分析方法,景观、森林、土壤空间异质性与格局分析,水生生物的种群空间格局分析,昆虫种群空间格局的动态等。大部分实例是作者这几年实际工作的结果。由于地统计学在国内生态学中的应用刚刚开始,作者所列举的实例难免有些浮浅或粗糙,因此在每一章节后面给出有关参考文献,请读者根据需要去阅读。

参考文献

- 王仁铎, 胡光道. 1987. 线性地质统计学. 北京: 地质出版社
- 李哈滨, 伍业刚. 1992. 景观生态学数量研究方法. 刘建国主编. 当代生态学博论. 北京: 中国科学技术出版社, 209—230
- 沈思渊. 1989. 土壤空间变异研究中地统计学的应用及其展望, 土壤研究进展, 2: 11—24
- 侯景儒, 郭光裕. 1993. 矿床统计预测及地质统计学的理论与应用, 北京: 冶金工业出版社
- Burgess T. M. and R. Webster. 1980a. Optimal interpolation and isarithmic mapping of soil properties. 1. The semivariogram and punctual kriging. J. Soil Sci., 31: 315—331
- Burgess T. M. and R. Webster. 1980b. Optimal interpolation and isarithmic mapping of soil properties. 2. Block kriging. J. Soil Sci., 31: 333—341
- Burgess T. M., R. Webster, and A. B. McBratney. 1981. Optimal interpolation and isarithmic mapping of soil properties. 3. Sampling strategy. J. Soil Sci., 32: 643—659
- Burgess T. M. and R. Webster. 1984. Optimal sampling strategies for mapping soil types. 1. Distribution of boundary spacings. J. Soil Sci., 35: 641—654

- Burrough P. A. 1983a. Multiscale sources of spatial variation in soil. I. The application of fractal concepts to nested levels of soil variation. *J. Soil Sci.*, 34:577—597
- Burrough P. A. 1983b. Multiscale sources of spatial variation in soil. II. A non-Brownian fractal model and its application in soil survey. *J. Soil Sci.*, 34:599—560
- Burrough P. A. 1987. Spatial aspects of ecological data. In *Data analysis in community and landscape ecology*. (R. H. Jongman et al ed). Pudoc Wageningen, The Netherlands
- Campbell J. B. 1978. Spatial variation of sand content and pH within single contiguous delineations. of tow soil mapping units. *Soil Sci Soc. Am. J.*, 42:460—464
- Clark I. 1979. *Practical geostatistics*. Applied Sci. Publishers LTO, London
- Cressie N. A. C. 1991. *Statistics for spatial data*. John Wiley & Sons, N. Y.
- David M. 1977. *Geostatistical ore reserve estimation*, Elsevier.
- Dutilleul P. and P. Legendre. 1993. Spatial heterogeneity against heteroscedasticity: an ecological paradigm versus a statistical concept, *Oikos*, 66:152—167
- Franklin J. F. and R. T. T. Forman. 1987. Creating landscape patterns by forest cutting: ecological consequences and principles. *Landscape Ecology*, 1:5—18
- Goodall D. W. 1952. Quantitative aspects of plant distribution. *Biol. Rev.*, 27:194—245
- Greig-Smith P. 1957. *Quantitative plant ecology*. Butterworth & Co. London
- Harlbert S. H. 1984. Pseudoreplication and the design of ecological field experiments. *Ecol. Monogr.*, 54:187—211
- Harvorson J. 1994. Geostatistics and ecology. *Bulletin of The Ecol. Soc. Am.*, 75:36—38
- Isaaks E. H. and R. M. Srivastava. 1989. *An introduction to applied geostatistics*. Oxford Univ. Press. New York
- Jakson R. B. and M. M. Cadwell. 1993. Geostatistical patterns of soil heterogeneity around individual perennial plants. *J. of Ecology*, 81:683—692
- Journel A. and Ch. Huijbregts. 1978. *Mining geostatistics*. Academic Press. London
- Jongman R. H. G. et al. 1987. *Data analysis in community and landscape ecology*. Pudoc. Wageningen
- Kareiva P. 1994. Space: The final frontier for ecological theory. *Ecology*, 76:1—1
- Katherine G. P. Kurts, and B. Andrew. 1995. Spatial variation in nitrogen availability in three successional plant communities. *J. of Ecology*, 83:357—367
- Kemp W. P., T. M. Kalaris, and W. F. Quimby. 1989. Rangeland grasshopper (Orthoptera: Acrididae) spatial variability: macroscale population assessment. *J. of Economic. Entomology*, 82:1270—1276
- Kreshaw K. A. 1973. *Quantitative and dynamic plant ecology* (2nd ed.). Elsevier. York. UK
- Leadley P. W., H. Li, B. Ostendorf, and J. F. Reynolds. 1996. Road—related disturbances in an arctic watershed; Analysis by a spatially explicit model of vegetation and ecosystem processes. In *landscape Function and Disturbance in Arctic Tundra* (ed by Reynolds et al.). Springer, New York. 387—415
- Legendre P. and M-J. Fortin. 1989. Spatial pattern and ecological analysis. *Vegentatio*, 80:107—138
- Levin S. A. 1992. The problem of pattern and scale in ecology. *Ecology*, 73:1943—1967
- Li H. and J. F. Reynolds. 1994. Ecological interpretation of semivariogram analysis: Quantification of spatial heterogeneity. Submitted to *Ecology*. in press
- Li H. and J. F. Reynolds. 1995. On definition and quantification of heterogeneity. *Oikos*, 73:280—284
- Matheron G. 1963. Principles of geostatistics. *Economic Geology*, 58:1246—1266
- McIntosh R. P. 1987. Pluralism in ecology, *Ann. Rev. Ecol. Syst.*, 18:321—341
- Orloci L. 1978. *Multivariate analysis in vegetation research* (2nd ed). The Hague, Junk
- Philips J. D. 1985. Measuring complexity of environmental gradients. *Vegetatio*, 6:95—102
- Pielou E. C. 1957. The effect of quadrat size on the estimation of the parameter of Neyman's and Thomas's distributions. *J. of Ecology*, 45:31—47
- Pielou E. C. 1959. The use of point-to-plant distances in the study of the pattern of plant populations. *J. of Ecology*, 47:607—613

- Robertson G. P. 1987. Geostatistics in ecology; interpolating with known variance. *Ecology*, 68:744—748
- Robertson G. P. , J. R. Crum, B. G. Ellis. 1993. The spatial variability of soil resource following long term disturbance. *Oecologia*, 96:451—456
- Robertson G. P. and D. W. Frekman. 1995. The spatial distribution of nematode trophic groups cross a cultivated ecosystem. *Ecology*, 76:1425—1432
- Robertson G. P. et al. 1997. Soil resources, microbial activity, and primary production across an agricultural ecosystem. *Ecol. Application*, 7:158—170
- Rossi R. E. , D. J. Mulla, A. G. Journel, and E. H. Franz. 1992. Geostatistical tools for modeling and interpreting ecological spatial dependence. *Ecol. Monogr.* , 62:277—314
- Schlesinger W. H. , J. A. Raikes, A. E. Hartley . 1996. On the spatial pattern of soil nutrients in desert ecosystems. *Ecology*, 77:364—374
- Schotzko D. J. and L. E. O' keetfe. 1989. Geostatistical description of the spatial distribution of *lygus hesperus* (Heteroptera; Miridae) in lentils. *J. Econ. Entomol.* , 82:1277—1288
- Sokala J. 1991. Ecological heterogeneity. Springer. New York
- Tilman D. 1984. Resource competition and community structure. Princeton Univ. Press. USA
- Tilman D. 1989. Plant strategies and the dynamics and structure of plant communities. Princeton Univ. Press. USA
- Turner M. G. and R. H. Gardner. 1991. Quantitative methods in landscape ecology. Springer-Verlag. New York
- Trangmar B. B. , R. S. Yost, and G. Uehara. 1985. Application of geostatistics to spatial studies of soil properties. *Advanced Agronomy*, 38:44—94
- Watt A. S. 1947. Pattern and process in the plant community. *J. of Ecology*, 35:1—22
- Webster R. 1985. Quantitative Spatial analysis of soil in the field. *Advance in Soil Science*, 3:1—70

第 2 章 样本数据的统计分析

- 总体、样本和随机样本
- 频数分布
- 统计特征数
- 两个样本之间的统计

§ 2.1 总体、样本和随机样本

§ 2.1.1 总体

总体(population)是指根据统计分析或研究的目的而确定的同类事物或现象的全体,有时也把具有共同性质的元素所组成的集合称为总体。有些书上也把总体称为母体。总体在数理统计中是一个非常重要的概念,因此有必要对总体的概念进一步说明。W. Cochran(1977)认为:第一,总体可以是指满足指定条件的元素或个体的集合。例如,一片森林可以作为一个总体,而每一株树木是其元素或个体。然而统计所感兴趣的并不是这些元素本身,而是这些元素的某种性质。例如我们的目的是要了解森林的蓄积量,因而统计上所感兴趣的是每株林木的树高(H)和胸径($D_{1.3}$)。因此总体又可理解为反映这些元素性质特征的一个集合。第二,总体的范围是根据研究目的来确定的,并不是固定不变的。仍以森林蓄积量为例,总体可以是 1 个树种的蓄积,也可以是林分中几个重要树种的蓄积。第三,总体可以分为目的总体和抽样总体。目的总体是指我们所希望元素的全体。但在抽样时要考虑怎样从总体中选取样本。因为这时的总体是比较抽象的,具体地确定总体的范围是困难的。例如要了解森林中某种害虫对树木的危害程度,我们要从发生虫害的树木群体中进行抽样,而“发生虫害的树木”就称为抽样总体。抽样的总体与目的总体有时一致,有时不一致。根据抽样的结果,只能对抽样总体作出有关推断的概率结论。而要把结论外推到目的总体上就必须从主观上进行判断。第四,总体可以分为定性和定量两大类,因此总体可分定性总体(又叫质量总体)和定量总体(数量总体)。第五,根据总体所包含的元素单位数是否有限,又可以分为有限总体和无限总体。在许多情况下,当总体单位虽为有限,但数目较多时,也可以把它看作一个无限总体。

§ 2.1.2 样本

研究的对象如果是总体,并要求得到总体的一些重要特征,但总体如果所包含的元素太多,往往不能逐一加以测定,因而,一般总是只能从总体中抽取若干个元素来研究。从总体中抽取若干个元素而构成的集合叫做样本,有时也称子样。我们的主要目的是从样本的观察或试验结果的特性来对总体特性作出统计估计与推断,因此要研究抽取样本的方法或技术,使样本能尽可能地反映总体的特性。样本抽选的方法有随机(或概率)抽样和

非随机(或非概率)抽样两种(Cochran 1977, Davis 1973),在数理统计中一般均采用随机抽样,这种抽取的样本的方法称随机样本。

§ 2.1.3 随机样本

在抽选样本时,总体的每一个元素被抽中的概率相同。这是基本的简单随机抽样。在具体进行抽样时,根据样本能否重复抽取分为重复抽样和不重复抽样两种。重复抽样是指每次抽取一个元素加以观察记录后再放回总体中去,参加下次抽选,直到抽足所需的样本数为止。显然一个元素可以被抽中若干次。不重复抽样是指已经抽出来的元素观察记录后,就不再放回总体中,每个元素能被抽中只有一次。总之,只有采用随机样本,样本才具有代表性,才能根据样本的特征去说明或推断总体的特征。

§ 2.1.4 总体和样本之间的关系

总体是客观存在的。譬如前面所述的一片森林的蓄积量,不论我们如何抽样,总体的性质和特征是不会改变的。但是总体中每个元素的数量表征是不同的,因而每一次试验(或观察)的结果也是不同的。如果在有限总体的情况下进行全面调查,或在无限总体的情况下进行无限次试验,我们会清楚总体各单位的数值分布的规律。在随机抽取的样本的条件下,样本的分布规律与总体的分布规律相似,即样本的分布在一定程度上反映了总体的分布。当样本的数量增大时,样本的分布规律接近总体的分布规律。显然当样本的数量增大到与总体相等时,两者的分布一样。

§ 2.2 频数分布

§ 2.2.1 观测数据的种类

任何试验或抽样调查所得的数据,根据所研究的性状、特征均可以分成两大类:即数量性状(或定量数据)和质量性状(或定性数据)。

§ 2.2.1.1 数量性状数据

数量性状的数据因有计数和测量两种方式,其所得的数据不同,分为间断性数据和连续性数据。所谓间断性数据,有时也称不连续变量,是指用计数方法获得的数据,以整数表示,在两个相邻的整数间不容许有小数的存在。例如林木株数、动物数量等。由于两个相邻的整数间是不连续的,因此称不连续或间断性数据。连续性数据是指用称量、度量或测量方法得到的数据,各个观察值并不限于整数,在两个相邻的数值间可以有小数存在。例如土壤 pH 值、土壤中氮元素的含量、树高等,其小数位数的多少,因称量精度而确定。这种数据称为连续性数据,或连续变量。

§ 2.2.1.2 质量性状数据

质量性状数据是指能观察但不能测量的性状数据资料。要从这类性状获得定量资

料,可以采用下列两种方法。首先应用统计次数的方法。在一定的总体或样本内,统计其具有某种属性的个体数量及不具备该种属性的个体数量,按这两种类型分别计算次数(或频数)或相对次数。这类数据也称为次数(或频数)数据。另外,是给予每个属性以相当数值的方法。例如土壤的肥力可分为好、中、差三级,分别给予 3, 2, 1 这样的数值等,这样就将定性资料转变为定量数据资料。

§ 2.2.2 频数分布表

在试验或抽样中,常常包含很多观测值,有时几百个或上千个。Davis(1973)认为,如果不对这些观测值进行整理,很难得到试验或样本所反映的明确的特征。如果把这些观测值按大小进行分组,制成频数分布表,就可以明显地看到观测数据的集中性和变异性的情况,从而对样本就有一个明确的了解。频数分布表的制作方法因数据种类不同而略有不同。

§ 2.2.2.1 间断性数据频数分布表

现在以森林更新样带调查资料为例,在小兴安岭天然阔叶红松林中设置 12 条样带,每条样带上相隔 100m 设置一块样地,调查红松幼树的株数。未加整理的资料见表 2.1。总共 144 块样地,红松幼树在每块样地上的株数均为整数(即间断性数据)。每块样地中红松幼树的株数在 0 ~ 31 的范围内变动,把所有调查数据按每样地的株数多少加以归类共分 12 组,组与组的相差为 2,称为组距。这样可得表 2.2 形式的频数分布表。从表 2.2 中看到,一堆杂乱的原始数据,经初步整理后,就可了解数据总体的基本特征,如在 144 块样地中有 2 株红松幼树的样地占 50%,有 4 株红松更新幼树的样地占 13% 等。

表 2.1 红松更新幼树数据

样带 样地	1	2	3	4	5	6	7	8	9	10	11	12
1	0	0	8	3	1	0	0	0	2	8	0	17
2	6	2	4	23	0	11	1	13	0	2	5	6
3	1	1	2	13	2	6	1	4	9	3	8	0
4	8	3	6	0	0	6	1	0	8	5	10	0
5	0	3	6	0	0	4	0	17	6	3	8	1
6	1	2	4	5	4	4	1	21	5	5	0	2
7	2	2	6	2	2	12	5	6	0	31	6	2
8	1	0	3	2	10	22	3	0	0	0	0	1
9	0	1	0	2	11	19	4	1	0	8	2	0
10	1	21	5	0	7	7	8	1	4	0	1	1
11	4	1	2	1	3	2	7	3	5	2	3	0
12	0	1	6	1	2	5	0	0	5	3	4	2

表 2.2 红松更新幼树频数分布表

频数			相对频数		
频数	相对频数		频数	相对频数	
2	75	0.521	14	3	0.021
4	19	0.132	18	2	0.014
6	21	0.146	20	1	0.007
8	12	0.083	22	3	0.021
10	3	0.021	24	1	0.007
12	3	0.021	32	1	0.007

§ 2.2.2.2 连续性数据频数分布表

连续性数据不可能按整数归组的方法来进行整理,首先必须确定组数、组距、各组的上下限,然后再按观测值的大小归组。表 2.3 是我国东北小兴安岭红松阔叶林下 100 个样方中土壤表层有机质的含量的原始数据。最大值为 3.90,最小值为 1.00。100 个样本数据分成 15 组左右比较合适,组距为 0.20。表 2.4 是连续性数据频数分布表。从表中的频数分布和相对频数分布得知,将近 30% 样方中土壤表层有机质含量在 2.00 ~ 2.40 之间。

表 2.3 100 个样方中土壤表层有机质含量(%) 数据

1.20	2.30	1.10	1.00	1.00	1.00	2.00	2.20	1.00	2.00
3.90	1.80	1.90	2.10	2.26	1.00	2.10	2.20	2.00	1.00
1.40	2.30	1.90	2.95	2.96	2.40	2.20	2.30	2.00	1.00
1.20	2.70	1.37	3.10	3.13	2.96	2.70	2.50	1.50	2.00
2.00	1.80	2.31	3.52	3.33	3.45	3.14	2.70	2.50	2.30
1.70	2.86	1.54	3.16	3.24	3.13	1.21	3.10	1.80	2.10
1.20	2.10	2.60	3.13	3.45	3.14	2.36	2.10	2.70	2.50
2.00	1.80	2.40	2.14	3.15	3.12	2.10	2.10	2.20	2.10
1.40	2.30	2.10	2.10	2.65	1.20	2.60	2.80	2.52	3.16
1.55	1.50	2.60	1.30	3.14	1.30	2.10	2.70	2.13	3.12

表 2.4 土壤表层有机质含量频数分布表

分组数列	频数	相对频数	分组数列	频数	相对频数
0.0 ~ 1.0	7	0.07	2.4 ~ 2.6	6	0.06
1.0 ~ 1.2	5	0.05	2.6 ~ 2.8	8	0.08
1.2 ~ 1.4	6	0.06	2.8 ~ 3.0	4	0.04
1.4 ~ 1.6	4	0.04	3.0 ~ 3.2	13	0.13
1.6 ~ 1.8	5	0.05	3.2 ~ 3.4	2	0.02
1.8 ~ 2.0	9	0.09	3.4 ~ 3.6	3	0.03
2.0 ~ 2.2	17	0.17	3.6 ~ 3.8	1	0.01
2.2 ~ 2.4	10	0.10			

在编制连续性频数分布表时,要注意组数和组距的确定。组距愈小,组数愈多。在确定组数时必须考虑到数据整理的目的。一方面在于简化数据资料,从中得到某种规律。另

一方面也要考虑到简化数据后能否真实地反映总体特征。如果组数过少,则频数分布表反映的精度较差,组数过多,数据过于分散,看不出数据资料的集中性特征,而且不便于进一步分析和计算。样本大小与组数的关系可参照表 2.5 来确定。

表 2.5 样本大小与组数的关系

样本数	组数	样本数	组数
50	5 ~ 10	300	12 ~ 25
100	8 ~ 18	500	15 ~ 30
200	10 ~ 20	1000	25 ~ 40

§ 2.2.2.3 定性数据频数分布表

对定性数据资料,也可以用类似频数分布的方法来整理。在整理前,把数据资料按各属性特征进行分类,分类数等于组数,然后根据各元素在属性上的具体表现,分别归入相应的组中,即可以得到定性数据频数分布表。

§ 2.2.3 频数分布图

试验数据或抽样的样本数据除了用频数分布表来表示外,也可用图形来直观地表示。频数分布图比频数分布表更直观地说明频数分布的情况。最常用的频数分布图是直方图(histogram),此外还有多边形图和条形图。直方图和多边形图均可用于间断性变量和连续性数据的描述,而条形图主要用于定性数据的描述。频数分布直方图一般是根据频数分布表来绘制(Tukey 1977)。现仍以表 2.4 为例,绘制森林土壤表层有机质含量的频数分布直方图。横坐标为组数,纵坐标为频数或相对频数,结果见图 2.1。

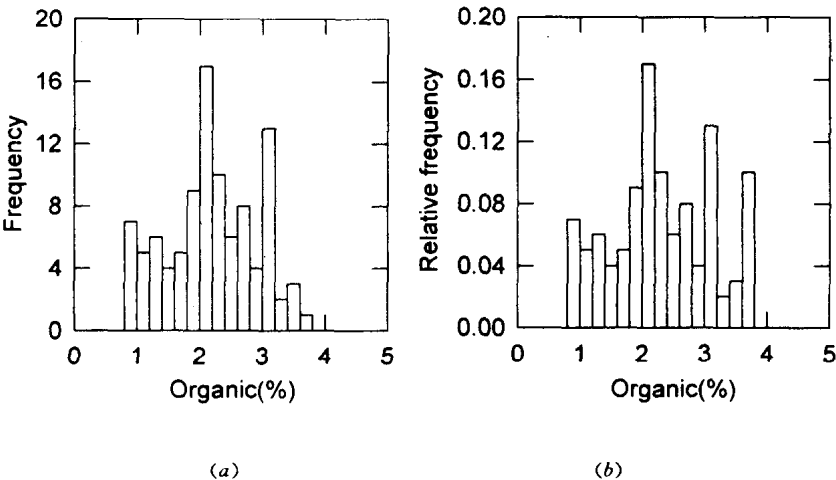


图 2.1 森林土壤表层有机质含量频数分布直方图(a)和相对频数直方图(b)