

[英] J·艾奇逊著

成分数据的统计分析

周 蒂 张光前 王仁铎 金友渔 魏 民 译

中国地质大学出版社

成分数据的统计分析

J. 艾奇逊 著

周 蒂、张光前、王仁铎、金友渔、魏 民 译

赵鹏大、周 蒂 审校

中国地质大学出版社

1989

内 容 简 介

本书是系统地介绍艾奇逊的成分数据的对数比统计分析方法的第一部专著。书中从介绍传统统计方法不适用于成分数据的原因开始,提出用对数比统计方法及逻辑正态分布为依据来研究成分数据,并详细介绍了这套方法和理论解决一系列实际问题的实施步骤;书中每章之后均附有思考、练习题,并从地学、生物学、医学、教育学等领域选取了40个实例来证明其理论和 方法,每例均附有实际数据以供读者分析、验证。介绍思想方法及应用实例的部分与介绍数学推导的部分分别属于不同的章节,以满足数学基础不同的读者的需要。

The Statistical Analysis of Compositional Data

J. Aitchison

First published in 1986 by

Chapman and Hall Ltd

ISBN 0-412-28060-4

成 分 数 据 的 统 计 分 析

J.艾奇逊 著

周 蒂等译

责任编辑 李继英

中国地质大学出版社出版

(武汉市喻家山)

湖南省地质测绘印刷厂印装 湖北省新华书店经销

开本 787×1032 1/16 印张 16.375 字数 419 千字

1990年6月第1版 1990年6月第1次印刷

印数1—1000册

ISBN 7-5625-0391-5/P·111

定价: 3.35元

译者的话

成分数据在社会科学和自然科学中真是太常见了！国民经济各部门的财政开支比例、大学毕业生分配方案、血液成分、药剂配方、以及地质学中的岩石化学全分析结果、岩石的矿物组成、沉积物粒度数据、化石种属比例等等。成分数据的研究对人类认识社会、认识自然至关重要。

成分数据常常具有高维数，因此研究者们往往借助于统计学，这方面的文章和专著屡见不鲜。遗憾的是，没有很多人注意到，成分数据是分布在有限区域内的带“定和”约束条件的数据，这些特点带来了闭合效应及参数分布上的许多困难。不认识这些困难，盲目地套用为无约束条件数据而设计的经典统计方法来研究成分数据是完全错误的。一些应用统计工作者，尤其是一些优秀的数学地质学家注意到这一问题，多年来为此大声疾呼（如Chayes, 1960, 1962, 1983; Krumbein, 1962; Samanov and Vistelius, 1959; Miesch, 1969; Butler, 1976, 1978, 1979, 1980, 1981, 1982）。但他们也没找到解决这些问题的真正途径。经典统计方法对成分数据的滥用仍然继续进行。

英国统计学家艾奇逊从70年代就开始研究成分数据统计分析中的问题。他认为如同带约束条件的方向数据有一套独特的统计分析方法一样，成分数据也应该有自己的一套统计方法。根据成分分量的比值不受闭合的影响、比值的对数常常服从正态分布的特点，他提出用成分分量比值的对数（简称“对数比”）来作为研究成分数据的基本单元，从而发展了一整套适用于成分数据的统计分析方法，解决了应用统计学中的一大难题。由于他在1982年发表的论文“The statistical analysis of compositional data”有很高的被引用率，他被授予英国皇家统计学会1988年研究奖章；这显示了统计学界对他的贡献的承认。

本书是系统地介绍艾奇逊的成分数据统计分析方法的第一本专著。正如第一章题头引语所示，艾奇逊赞赏“问题的答案不仅取决于观测数据，而且取决于问题本身”的观点。他在本书中首先针对各个应用领域中的成分数据提出一系列有实际意义的研究问题，然后在各章中逐渐展示解决这些问题的正确办法，章末还提出一些思考、练习题。这种写作方式不仅使得复杂的数学问题变得浅显易懂，而且对读者将书中提出的统计方法应用于各自的实践很有启发。数学关系的推导集中在第五章，以便有兴趣的读者阅读。与专著出版的同时作者还发表了适用于IBM/PC系列微机的程序集CODA，以帮助读者在微机上实现成分数据的统计分析。这一切，使得本书不仅是一部提出新的统计方法的专著，而且可用作教科书或实际应用统计学家的参考书。

本书的翻译工作由下列同志共同完成：周蒂，前言、第1、2、3章、关键词中英文对照；王仁铎，第4、5、13章；张光前，第8、9、10、11、12、14章；金友渔，第6、7章；魏民，附录A、B、C、D。最后由周蒂统稿，赵鹏大教授和周蒂审校。错误疏漏之处，敬请读者批评指正。程序集CODA的软盘和使用说明以后将分别推出。

1988年10月6日

目 录

前 言	(1)
第一章 成分数据——一些挑战性的问题	(3)
1.1 绪 论	(3)
1.2 岩石的地球化学成分	(3)
1.3 不同深度的沉积物	(5)
1.4 三元图	(5)
1.5 部分分析及子成分	(6)
1.6 管理态度	(8)
1.7 家庭预算调查	(9)
1.8 成人和儿童的类固醇代谢模式	(9)
1.9 某统计学家的活动模式	(10)
1.10 白细胞成分的标定	(10)
1.11 水果评价	(11)
1.12 焰火混合物	(11)
1.13 蛤的生态	(11)
1.14 文献注记	(12)
练习题	(12)
第二章 作为样本空间的单形	(15)
2.1 样本空间的选择	(15)
2.2 成分和单形	(15)
2.3 空间、向量、矩阵	(17)
2.4 基和成分	(18)
2.5 子成分	(19)
2.6 合 并	(21)
2.7 分 割	(22)
2.8 摄 动	(24)
2.9 成分数据的几何表示法	(25)
2.10 文献注记	(26)
练习题	(26)
第三章 成分数据分析的特殊困难	(28)
3.1 绪 言	(28)
3.2 高维数问题	(29)
3.3 缺乏可解释的协方差结构	(30)
3.4 参数建模的困难	(33)
3.5 混合变化的困难	(35)
3.6 文献注记	(35)
练习题	(36)

第四章 协方差结构	(37)
4.1 基本原理	(37)
4.2 协方差结构的详细说明	(38)
4.3 成分变异数组	(39)
4.4 根据原始均值向量和协方差矩阵来恢复成分变异数组	(41)
4.5 子成分分析	(42)
4.6 协方差结构的矩阵明细表	(43)
4.7 一些重要的初等矩阵	(47)
4.8 矩阵明细表间的关系	(47)
4.9 香岩成分的矩阵估计	(48)
4.10 对数比和对数衬度 (logcontrasts)	(49)
4.11 基的协方差结构	(50)
4.12 评 述	(51)
4.13 文献注记	(51)
练习题	(51)
第五章 矩阵协方差明细表的性质	(53)
5.1 对数比记号	(53)
5.2 对数衬度的方差和协方差	(53)
5.3 置 换	(54)
5.4 P 和 Q_P 矩阵的性质	(55)
5.5 涉及 Σ 置换不变量	(55)
5.6 协方差矩阵的逆	(57)
5.7 子成分	(58)
5.8 T, Σ, Γ 的特征等价性	(59)
5.9 对数比不相关成分	(61)
5.10 各向同性协方差结构	(63)
5.11 文献注记	(65)
练习题	(65)
第六章 单形空间上的逻辑正态分布	(66)
6.1 引 言	(66)
6.2 加法逻辑正态分布	(66)
6.3 密度函数	(67)
6.4 矩性质	(68)
6.5 对数正态基的成分	(69)
6.6 类保留性质	(69)
6.7 条件子成分的性质	(72)
6.8 扰动的性质	(73)
6.9 中心极限定理	(73)
6.10 对数衬度的特征	(74)
6.11 与狄利克雷分布类的关系	(74)

6.12 统计分析的可能性	(76)
6.13 乘法逻辑正态分布类	(77)
6.14 分割逻辑正态分布类	(78)
6.15 某些符号	(80)
6.16 文献注记	(81)
练习题	(82)
第七章 成分的对数比分析	(84)
7.1 绪 言	(84)
7.2 μ 和 Σ 的估计	(85)
7.3 有效性: 逻辑正态性检验	(85)
7.4 假设检验的对策和技术	(89)
7.5 关于 μ 和 Σ 的假设检验	(91)
7.6 对数比线性模拟	(95)
7.7 对数比线性假设的检验	(97)
7.8 关于对数比线性模拟的进一步探讨	(100)
7.9 对数比线性模拟的应用	(102)
7.10 预测分布、非典型指数和异点	(104)
7.11 统计判别	(106)
7.12 条件成分模拟	(109)
7.13 文献注记	(110)
练习题	(110)
第八章 降维技术	(111)
8.1 引 言	(111)
8.2 原始的主分量分析	(112)
8.3 对数衬度主分量分析	(113)
8.4 对数衬度主分量分析的应用	(116)
8.5 子成分分析	(117)
8.6 子成分分析的应用	(119)
8.7 典型分量分析	(121)
8.8 文献注记	(123)
练习题	(123)
第九章 基和成分	(124)
9.1 基本原理	(124)
9.2 协方差关系	(124)
9.3 主分量与典型分量的对比	(127)
9.4 分布的相互关系	(130)
9.5 成分的不变性	(131)
9.6 家庭预算分析方面的应用	(133)
9.7 在临床生物化学方面的应用	(135)
9.8 对过去的一个形态和容量分析的重新评价	(135)

9.9 文献注记	(137)
练习题	(138)
第十章 子成分和分割	(139)
10.1 引言	(139)
10.2 完全的子成分独立性	(139)
10.3 一阶分割	(143)
10.4 分割的有序序列	(146)
10.5 防止误解的说明	(149)
10.6 高阶分割	(149)
10.7 文献注记	(150)
练习题	(151)
第十一章 不规则成分数据	(152)
11.1 引言	(152)
11.2 成分的误差建模	(152)
11.3 影响精度原因的分析	(154)
11.4 不精确度和独立性检验	(156)
11.5 舍入零或微迹零	(157)
11.6 真 零	(160)
11.7 缺失分量	(162)
11.8 文献注记	(163)
练习题	(164)
第十二章 协变作用中的成分	(166)
12.1 引言	(166)
12.2 标 定	(166)
12.3 处理前后的问题	(167)
12.4 混合试验	(168)
12.5 焰火配料上的应用	(171)
12.6 成分分类	(172)
12.7 地质分类中的应用	(175)
12.8 文献注记	(177)
练习题	(178)
第十三章 单形上的另外一些分布	(179)
13.1 狄利克雷类的某些推广	(179)
13.2 逻辑正态类的某些推广	(180)
13.3 摘要说明	(182)
13.4 $\mathcal{A}^d(\alpha, B)$ 类	(182)
13.5 极大似然估计	(184)
13.6 中性和分割独立性	(186)
13.7 子成分独立性	(188)
13.8 具有成分不变性的广义对数正态伽马分布	(189)

13.9 讨 论	(190)
13.10 文献注记	(191)
练习题	(191)
第十四章 其他问题	(192)
14.1 引 言	(192)
14.2 多向成分	(192)
14.3 多阶段成分	(194)
14.4 复合成分	(195)
14.5 成分数据的核密度估计	(197)
14.6 成分的随机过程	(199)
14.7 与贝叶斯统计分析的关系	(200)
14.8 成分数据和方向数据	(201)
练习题	(202)
附 录	(203)
A 初等矩阵的代数性质	(203)
B 文献目录	(204)
C 成分数据分析的计算机软件	(213)
D 数据集	(214)
关键词中英文对照	(239)

前 言

早在1897年, Karl Pearson在一篇讨论伪相关的权威性文章中第一次指出: 试图解释在分母及分子中含有公共部分的那些比例之间的相关性是很危险的。这意味着, 研究作为整体的各部分的相互关系的成分数据分析将充满困难。历史证明他是对的; 在以后的年代里, 可以说直到今天, 还没有哪个数据分析领域象成分数据的统计分析那样充满了概念的混淆及方法的滥用。

成分数据的一个特殊的内在特征是, 其组分服从总和为1的限制条件。对于其他形式的带限制条件的数据, 如方向数据, 其每一方向矢量服从长度为1的限制条件, 科学家和统计学家很快就认识到需要采用专门的统计方法; 而至今已经有了大量关于成功的方向数据统计分析的文献。令人费解的是, 成分数据的限制条件更为简单, 但类似的认识及方法研究却进展缓慢。在实际的成分数据分析中, 定和限制常常被有意或无意地忽略, 为不带限制条件的数据而设计的“传统”统计方法被不适当地滥用, 造成了灾难性的恶果。也有一些科学家和统计学家认识到成分数据统计分析中的困难, 但似乎他们的大量精力不是用于创造新的方法, 而是耗费于警告同行们传统的概念对于成分数据是解释不通的, 以及越来越详细地从各个方面阐明传统方法所导致的谬误。

本书的目的在于清晰而完整地介绍近年来专为成分数据而设计的一套新的统计方法。按照应用数学的写作方式, 我们将从多种领域内的成分数据分析的实际问题出发引入统计学的概念、原则及方法, 又反过来通过应用于成分数据而进一步阐明这些方法。强调提供有效而可行的方法, 而不是进行详细的证明和推导。任何关于统计学的某个新的重要领域的第一部著作的一个基本任务是使读者充分熟悉其统计方法, 并有足够的信心将这些方法应用于陌生的成分数据。为了有助于确立这种信心, 在每章的结尾为读者提供了一系列练习题。

成分数据的分析者们将认识到, 数据的多元形式要求统计分析方法也是多元的。因此, 要求本书的读者不仅熟悉统计推断的基本知识, 而且对传统的多元统计分析有一定的、哪怕是极少的了解。对于曾经应用传统多元方法来研究成分数据(尽管这种应用是错误的) 的人, 接受本书中提供的方法将毫无困难。适应于定和数据的新方法的美妙之处在于它的简单易行。

为了有效地运用成分数据统计分析中的许多特殊方法, 设计了一个方便用户的对话式微机统计程序包CODA。附录C中提供了有关CODA的技术细节及其软盘和手册的索取办法。

本书得到了许多人的帮助。香港大学的一些同事在成分数据统计分析的研究中给予的合作使完成本书成为可能。感谢在澳大利亚、加拿大、新西兰、中国、英国、美国举行的会议或讲座中的讨论者们, 尤其是1982年元月在伦敦皇家统计学会参加论文讨论的学者们。普林斯顿大学的研究生们在1981年秋季以及香港大学的研究生们在1984—1985学年的课程中提出了许多很好的问题, 感谢罗露茜女士耐心而幽默地完成了本书的文字处理工作, 提供了完美的打印稿。感谢我的妻子Muriel在炎热的香港的夏日, 为本书的脱稿牺牲了度假。

参考系统

定义、表格、插图、性质（定理）、方程、问题和练习题等的编号是在章、节、顺序号之间加脚点，例如定义2.1，表1.6，图6.2，性质5.3，问题1.2.4。方程的编号写在圆括号中，放在右侧，例如（4.2）表示第四章的第2个方程。

若干数据集在书中被反复引用，以帮助概念的引入和方法的建立，并说明方法的应用及向读者提出新的问题。为了参考的方便，这些数据集均在书末附录D中，而在书中被称为数据1、数据2等等。

为了避免打断叙述过程，参考文献将大多列入每章最后的文献注记中。正文中的参考文献仅注明作者及发表年代。

基本矩阵性质列入附录A中，在正文中以F1、F2、G1、H1等符号代表。

第一章 成分数据——一些挑战性的问题

答案不仅取决于观测数据，而且还取决于问题本身——这句话有时被当作谬论，却实在应该是常识。

Harold Jeffreys: 概率论

1.1 绪 论

如果向量由代表某整体的各部分的非负元素 $x_1, \dots, x_D$ 组成，它显然应服从单位和约束条件

$$x_1 + \dots + x_D = 1, \quad (1.1)$$

由这样一些向量组成的数据集被称为成分数据，它在许多领域内起重要作用。成分数据的各向量间常呈现显著的变异性，因而必须利用统计分析的手段来进行研究和解释。然而，十分常见的是忽略了单位和约束条件(1.1)，或将约束条件不恰当地结合于统计建模之中，导致了不充分或不合适的分析，以及不可靠或被歪曲的推断。本书的唯一目的是建立一套简单的、适用于成分数据的统计方法，并阐明它们的实际用途。

作为开始，我们将在本章介绍从若干不同领域中选取的各种成分数据分析问题。这样不仅可展现问题的广度，而且可引出以后各章中所介绍的概念和方法。发展概念和方法的第一步就是鉴别出适合于每种特殊情况的一些问题，并将这些问题用公式表示出来。

在某些应用领域内，原始数据量很大。为了给读者提供问题的某些形式，我们不得不将数据集压缩成教科书可容纳的体积。在这样做时，我们注意了保留每个问题的基本性质以及提供该分析所需的全部方法。由于书中多处要提到这个或那个数据集，我们将全部数据集列在书后的附录D之中，以便查找。这些数据集被按顺序编号，并简称为“数据1”、“数据2”等等。

1.2 岩石的地球化学成分

岩石地球化学成分的统计分析是岩石学的基本问题之一，这种成分通常表示为10种或更多种主要氧化物的重量百分比。例如，表1.1中的成分为6个标本的10种主要氧化物的重量百分比，其中3个为二叠纪的标本，3个为二叠纪后的标本，均选自Carr(1981)所研究的102个标本之中。变化不仅存在于不同种类的标本之间，也存在于同类的不同标本之间；必须通过某种形式的统计分析才能充分认识二叠纪和二叠纪后岩石的主要氧化物成分的模式。

为了通过更小的数据集来研究这类问题的特性，我们在数据1和2中给出了50个岩石标本，其中25个属香岩（数据1）、25个属港岩（数据2）。每种成分包含5种矿物（阿拜矿、布兰带矿、柯奈矿、多拜矿、恩带矿）的重量百分比；在以后的讨论中将这5种矿物简称为A、B、C、D、E。（香岩、港岩、5种矿物、以及后面将要出现的波克岩、柯克岩等均为虚构的名

表1.1 二叠纪和二叠纪后岩石的某些类型的主要氧化物成分

类 型	主 要 氧 化 物 重 量 百 分 比									
	SiO ₂	TiO ₂	Al ₂ O ₃	全铁	MnO	MgO	CaO	Na ₂ O	K ₂ O	P ₂ O ₅
二叠纪	60.54	1.32	15.22	6.95	0.21	2.33	3.18	4.81	4.84	0.60
	54.30	1.24	16.67	8.70	0.07	4.24	8.34	3.41	2.52	0.49
	52.17	0.82	20.05	8.38	0.10	2.28	9.29	3.22	2.99	0.69
二叠纪后	55.95	1.26	18.54	7.24	0.28	1.20	3.30	6.14	5.67	0.45
	45.40	1.34	20.14	8.00	0.06	9.29	9.59	3.89	1.38	0.90
	46.59	1.06	15.99	11.20	0.30	10.50	10.45	2.03	1.45	0.43

称——译者注)。

每一数据集都呈现出成分数据集的特征:

1. 数据中的每一行代表一块岩石样品,或更一般化地为一个样品、一次试验、或一个测量单位;

2. 每一列对应于一种矿物,或更一般化地称为成分中的一个组分或元;

3. 每个元素都是非负的;

4. 每行的元素之和为1,亦即百分之百。

这样,位于第 r 行和第 i 列的元素 x_{ri} 代表了第 i 种组分(或元)在第 r 个样品中的比例(或分量)。对于数据1和2中的由5个元组成的各25个样品, r 由1变到25, i 由1变到5。在考虑一般的成分数据阵列时,我们将假设它含有由 D 个元组成的 N 个样品,而行(或样品)的和为1的特性带来了以下的约束条件:

$$\sum_{i=1}^D x_{ri} = 1 \quad (r=1, \dots, N)。$$
 (1.2)

若将完整的数据阵列表示为 $N \times D$ 矩阵 X ,则约束条件(1.2)可表示成矩阵形式

$$Xj_D = j_D,$$
 (1.3)

其中 j_D 代表一个 $D \times 1$ 向量,其 D 个元素均为1。

为了发展适用于香岩和港岩数据集的统计方法,首先要提出一些典型的问题。

问 题 1.2

1. 如何令人满意地描述香岩成分的变异模式?
2. 有一块被声称香岩的新岩石标本,其(A、B、C、D、E)5个元的比例为(40.0、20.4、13.0、9.1、12.6),我们能说这块标本就成分来看是典型的香岩吗?如果不是,能否测量其不典型程度?
3. 对于某一特定的岩石类型,是否能定义一种协方差或相关结构,使我们能提出和检验关于该结构的各种有意义的假设?
4. 在单位和约束条件下可能有哪些形式的统计独立性?
5. 通过研究成分的一个子集(如A、B、C)中的各组分的相对数值,我们能对该成分的变异模式得出何种程度的认识?如何定量地描述用这种子成分分析方法来研究成分数据时信息量的损失大小?
6. 香岩和港岩的变异模式不同吗?如果不同,能否设计一种根据成分来进行分类的方便形式?
7. 能否判断根据选出的某些组分进行的分类将会与根据全成分进行的分类同样有效?

类似的问题也适用于数据3和4,它们分别给出了波克岩和柯克岩的25个标本的A、B、C、D、E 5种矿物的重量百分比。这两组数据还包括每一标本的采样深度,对于柯克岩还有孔隙度数据。以后我们将会提出

涉及这些伴随信息的一些问题，目前我们暂时集中精力考虑这两组成分的一个显著特点——它们的变异性较之香岩和港岩要小得多。

8. 是否能通过定义总变异性的某种测度来定量描述某一成分的变异性？

1.3 不同深度的沉积物

在沉积学中通常将沉积物分为3个互斥完备组分：砂、粉砂、粘土；它们的重量比例就是沉积物的成分。数据5中记录了在一个北极湖中不同水深处采取的39个沉积物样的成分。这里显然有以下两个问题：

问题 1.3

1. 沉积物的成分与水深有关吗？
2. 如果有关，如何将这种关系定量化？

对于这个成分数据集， $N = 39$ ， $D = 3$ ，成分数据矩阵 X 为 39×3 矩阵。水深被称为伴随变量或协变量，它有可能解释成分的某些变化。用熟悉的统计学术语来说，我们面临的是一个多元回归问题，沉积物成分是因变量，水深是自变量。但这一问题有一个特殊的方面：因变量向量的各分量满足单位和约束条件。

对数据3中波克岩的成分与水深的可能关系也可提出类似的问题。

1.4 三元图

三元成分（如数据5中的砂、粉砂、粘土）变异性的一种很方便的图示方法是三元图，亦称为参考三角形或重心坐标空间。三元图在许多领域内，尤其在地学领域内应用极为普遍。以下讨论三元图在沉积物成分研究中的应用。

图1.1中三角形的顶点分别为1、2、3，各边相等，各高都为1。在三角形内的任意点P到1、2、3的对边的垂线 x_1 、 x_2 、 x_3 满足条件

$$x_i \geq 0 \quad (i = 1, 2, 3),$$

$$x_1 + x_2 + x_3 = 1. \quad (1.4)$$

反之，满足条件(1.4)的任一向量 (x_1, x_2, x_3) 都必然对应于三角形123之内的一个到三边的垂线分别等于 x_1 、 x_2 和 x_3 的点。这种三元成分与三角形123内的点之间的一一对应关系使我们有了一个表示三元成分的简单方式。我们会注意到，在这种表达方式中，仅当代表点位于三角形123之内时式(1.4)中的不等式才严格成立； x_i 越大，代表点离顶点 i 的对边越远，或粗略地说离顶点 i 越近。如果一些三元成分中有两个组分 x_2 和 x_3 的比值为常数，它们的

的代表点将会落在一条通过顶点1的直线上。供绘制三元图用的三角坐标纸可在市场上买到。

利用这种表达形式，可将数据5中用百分比表示的砂-粉砂-粘土成分（如77.5，19.5，3.0）改写为小数形式（如0.775，0.195，0.030），然后表示成三元图1.2中的一个点。由

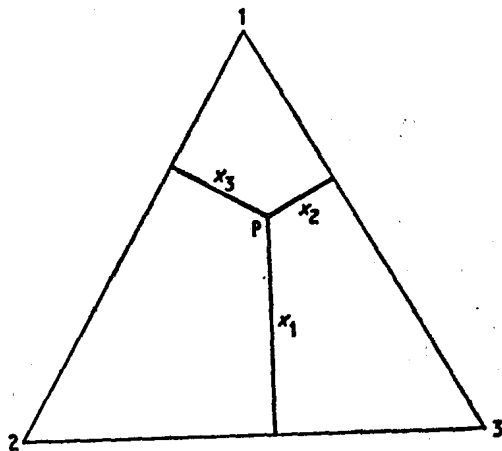


图1.1 三元成分 (x_1, x_2, x_3) 在参考三角形中的表示法

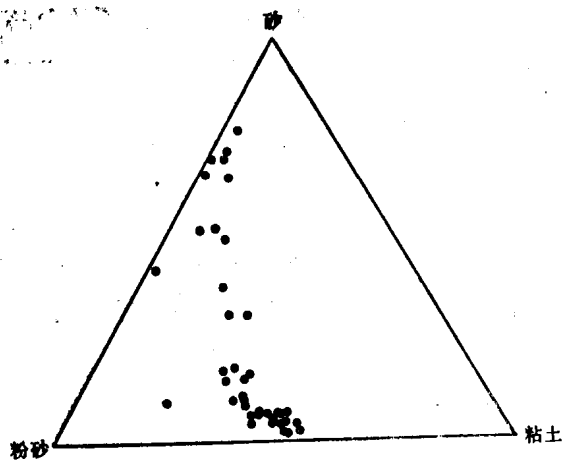


图1.2 一个北极湖中39个沉积物样品的（砂、粉砂、粘土）成分的三元图

图可见，相当一部分成分中砂的含量很低，而粉砂与粘土的含量大致成比例。两个组分间比值的变化大小可以用通过互补顶点的直线扫过全部代表点所需的角度大小来衡量。例如，对这组沉积物样品，砂/粘土比值的变化要比粉砂/粘土比值的变化大得多。

单靠三元图显然不能揭示沉积物成分对水深的依赖关系；为此需要作一个三维图示，如图1.3所示的三角柱图。将数据5中的各点投到相应水深处的三元图上，就可显示出数据的全部特征，我们可试图沿着这个三角柱观察代表点的分布是否有系统变化。尽管这类三维图

在数据分析上的作用是有限的。再者，这类图示也无法应用于多于三元的成分或多于一个伴随变量的情况。

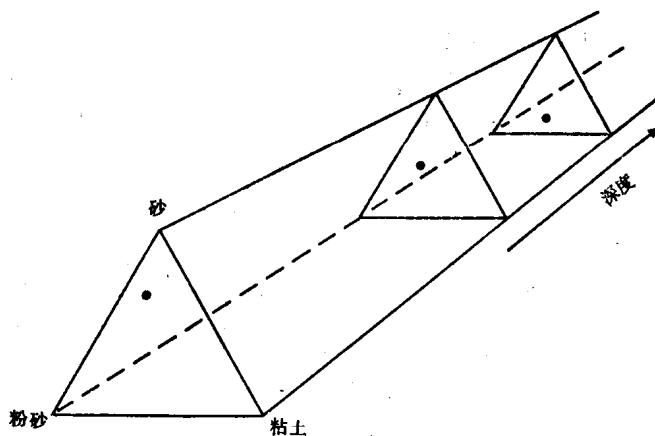


图1.3 沉积物成分和深度的三角柱图

1.5 部分分析及子成分

在提出关于更多成分数据集的问题之前，先让我们用三元图来阐明在问题1.2.5中提出的有关数据1和2中香岩和港岩的部分分析的形式。

在1.4节中我们已看到了怎样获得三元成分变异性的图示。二元成分的图示方法更为简单：用一条单位长的线段去联结点1和2（图1.4）；如果 x_1 较大，代表点P离端点2较近。

对于四元成分，我们必须进入三维空间，用一个正四面体1234来表示其变异性。与作为三元成分的参考格架的三角形123的作用相当，这个四面体的每个面上的高都等于1。在图1.5中，组分 x_i 对应于由代表点P向着与顶点*i*相对的三角形所作的垂线。还没有什么令人满意的方法能图示四元以上成分的变异性，因此人们倾向于进行部分分析，即着眼于该成分的某种子成分。

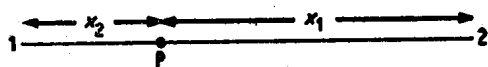


图1.4 二元成分（ x_1 、 x_2 ）在点1和2的连线上的表示法

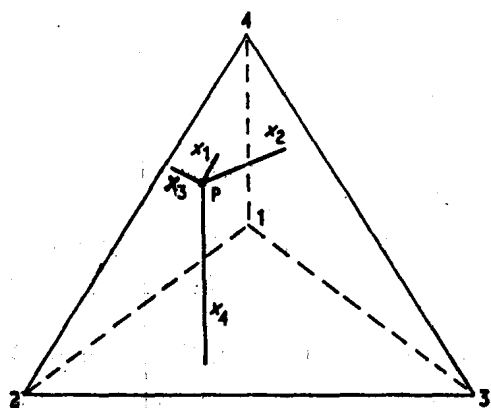


图1.5 四元成分 (x_1, x_2, x_3, x_4) 在参考四面体1234中的表示法

对于香岩矿物组成的任一子集, 如子集 A, B, C, 我们可通过重新标度来使各组分比例之和为1, 从而形成一个子成分。例如, 香岩的第一块标本 H1 的 ABC 子成分 (以百分比表示) 是

$$100 \times (48.8, 31.7, 3.8) / (48.8 + 31.7 + 3.8) = (57.9, 37.6, 4.5)。$$

香岩的25个ABC子成分的完全集见表1.2。每个这样的子成分实际上都是一个三元成分, 因而可以在三元图ABC上表示出来 (图1.6)。类似地, 香岩的子成分ADE也可在三元图ADE上表示出来 (图1.7)。

子成分ABC和ADE的变异模式有很大差别:

表1.2 数据1中25块香岩标本的ABC子成分

标本号	百分比			标本号	百分比			标本号	百分比		
	A	B	C		A	B	C		A	B	C
H1	57.9	37.6	4.5	H10	47.4	51.8	0.8	H19	53.9	20.1	26.0
H2	59.5	29.4	11.1	H11	61.8	24.1	14.1	H20	55.8	14.4	29.8
H3	46.1	11.3	42.6	H12	53.0	43.8	3.2	H21	57.8	21.0	21.2
H4	62.1	29.1	3.8	H13	40.8	10.6	48.6	H22	61.1	30.6	8.3
H5	51.8	44.8	3.4	H14	53.3	16.6	30.1	H23	57.1	40.0	2.9
H6	63.2	31.7	5.1	H15	53.1	22.0	19.9	H24	57.1	20.3	22.1
H7	54.3	40.1	5.6	H16	40.1	9.1	50.8	H25	57.0	30.9	12.0
H8	41.3	6.3	51.9	H17	48.8	50.1	1.1				
H9	51.3	14.7	33.5	H18	53.3	38.0	3.7				

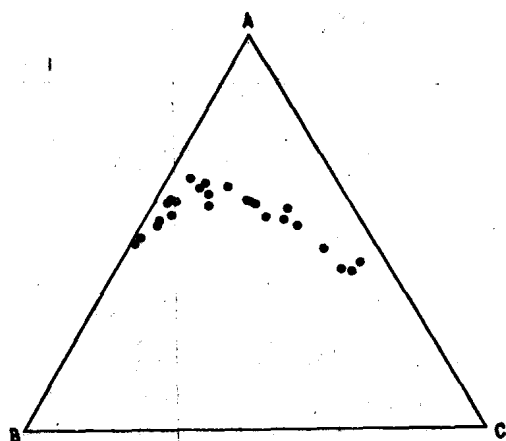


图1.6 25块香岩标本的ABC子成分

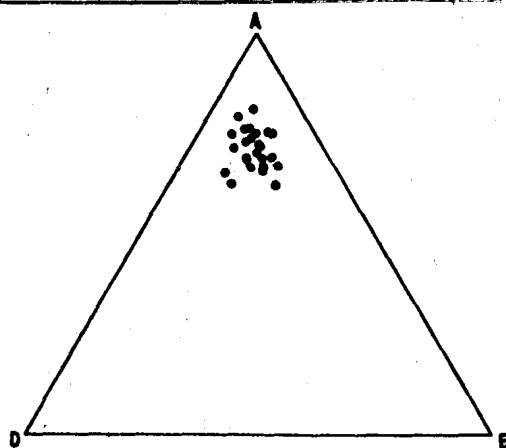


图1.7 25块香岩标本的ADE子成分

前者的样点分布呈香蕉状或飞碟状, B/C 比值的变化比其余两个比值的变化大得多; 而后的样点分布密集呈梨形, A, D, E 三个组分的任何比值都变化不大。

此时, 一系列难解的问题会出现在脑际:

问题 1.5

1. 是否应该将图1.6中的明显弯曲解释为香岩ABC子成分的某种变化趋势?
2. 在何种意义下ABC子成分能代表全成分变异性的基本模式?
3. 在这方面哪个子成分较好, ABC还是ADE? 如果我们判定ABC包含了全成分的大部分信息, 是否可能有它的某一个二元子成分, 如AB、AC、或BC, 也同样令人满意?
4. 能否判断图1.6所示的香岩ABC子成分与图1.8所示的港岩ABC子成分模式有无差别? 如果不能, 怎样才能选出一个三元子成分, 使它既能区别代表香岩和港岩的基本模式, 又能突出它们之间的差异?
5. 如果10个可能的三元子成分中没有一个是满足上述要求, 我们可以下结论说矿物成分对区别香岩和港岩不起作用吗? 更一般地说, 对于三元子成分, 还有没有成分数据的其它图示方法比三元图更令人满意?

地球化学研究中部分分析的一个常见例子是AFM三元图, 它由以下三种组分的相对比例组成:

A— $\text{Na}_2\text{O} + \text{K}_2\text{O}$, 即碱性组分;

F— Fe_2O_3 ;

M— MgO 。

作为一个典型例子, 在数据6中给出了无斑隐晶斯凯熔岩的23个AFM成分, 相应的三元图见图1.9。

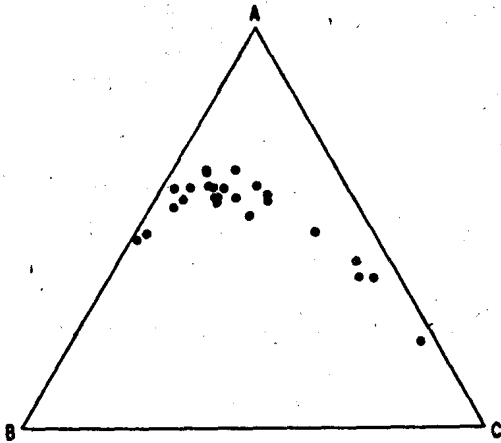


图1.8 25块港岩标本的ABC子成分

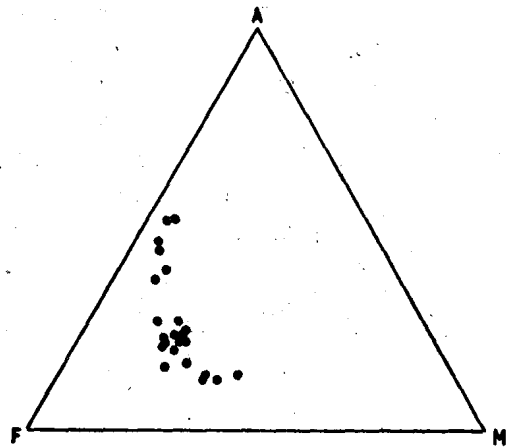


图1.9 23块无斑隐晶斯凯熔岩的AFM三元图

1.6 管理态度

迄今我们对成分数据所提的问题都是有关岩石和沉积物的地质问题。现在让我们进入另一个领域——人类行为的研究, 并提供一个日益令人关注的调查项目的简化形式。

数据7记录了对一位主管和三名属员之间关系的调查结果。主管对一名属员进行一小时技术指导为一次指导, 在十二周内每名属员接受了六次指导, 每两周一次。对总共十八次指导都作了记录, 主管的“语句”被分为4类:

C—命令, 给属员一条专门指示;

D—要求, 向属员提出一定的问题;