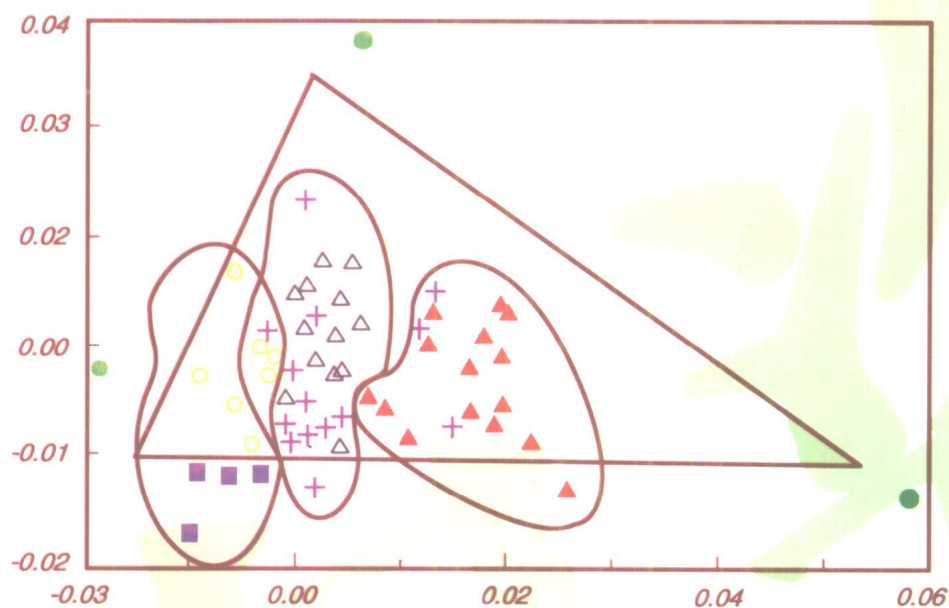


中国古陶瓷与多元统计分析

罗宏杰 编著

中国古陶瓷与多元统计分析



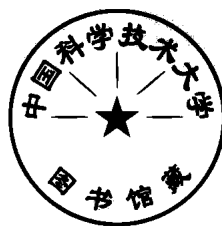
中国轻工业出版社

中国轻工业

3297
042
768

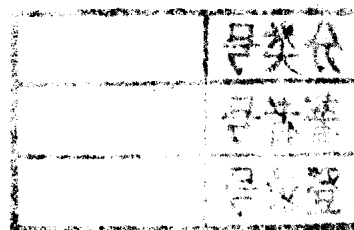
中国古陶瓷与多元统计分析

罗宏杰 编著



中国轻工业出版社

内 容 简 介



本书从多元统计分析的数学方法入手,侧重于对应分析、聚类分析以及判别分析等数学方法在中国古陶瓷研究中的应用,并向广大科研工作者提供了实用的多元统计分析使用软件及其软件的使用方法。同时本书收集了近几十年来古陶瓷研究的数千个中国古陶瓷胎、釉化学组成数据,并按古陶瓷种类、地区、窑口以及烧造时期等多个指标排序印刷,十分便于研究者多种目的的使用。

本书属陶瓷、考古以及现代数学等多种学科的交叉学科范畴,是被国外称为“计算机考古”学科中的一个分支。本书适合于大专院校陶瓷、考古专业师生及陶瓷学家、考古学家使用。

图书在版编目(CIP)数据

中国古陶瓷与多元统计分析/罗宏杰编著. -北京:中国轻工业出版社, 1997.10

ISBN 7-5019-2124-5

I. 中… II. 罗… III. 古代陶瓷-多元分析-中国 IV. K876.3

中国版本图书馆 CIP 数据核字 (97) 第 18585 号

1998.6.17
中国报道社
北京书刊发行部
No. 7986834

责任编辑:李春姪

*

中国轻工业出版社出版发行

(100740 北京市东长安街6号)

中国人民警官大学印刷厂印刷

新华书店经销

*

开本: 787×1092 1/16 印张: 15.25 字数: 353 千字

1997年10月第1版 第1次印刷

印数: 1—1000 定价: 25.00 元

序

中国古陶瓷既有悠久的历史，又有丰富的内涵，是中华民族光辉传统文化的重要组成部分。近代中外学者，特别是中国学者对它展开的科学技术研究，揭示了它的许多科学技术内涵，使人们了解到中国古陶瓷不仅是艺术百花园中的一朵奇葩，而且也是科技王国中的一座宝库。

罗宏杰博士自 80 年代以来，在攻读硕士学位和博士学位过程中以及随后在西北轻工业学院的教学和科研中，一直将现代科学理论和测试技术应用于中国古陶瓷的研究。特别是他在导师的指导下，将多元统计分析技术应用于中国古陶瓷的研究，与此同时并建立了国际上第一个中国古陶瓷化学成分数据库，使中国古陶瓷研究步入一个更深的层次和走上一个新的台阶。本书即是他多年来在这方面研究的成果和心得的总结。

《中国古陶瓷与多元统计分析》是国内外第一本阐述多元统计分析方法在中国古陶瓷研究中应用的专著。作者充分利用多元统计分析方法的功能，对中国古陶瓷万余年来从陶到瓷的发展过程加以详细地阐述和讨论，并对某些地区历代名瓷的特征及其工艺演变规律加以深入地分析和论证。这些结果将为弘扬中华传统文化并为历代名瓷的断源和断代以及它们的恢复生产提供科学依据。另外值得一提的是本书还简介了中国古陶瓷化学成分数据库，为广大古陶瓷研究者快速、准确、方便地提供了各种表达形式的中国古陶胎、釉化学组成的数据。

本书即将付印出版，作为第一读者，本人十分高兴。深觉全书结构严谨，内容系统，文字流畅，讲解清楚，具有鲜明的特色和很高的参考价值。相信本书的出版将会进一步推动中国古陶瓷科学技术研究更广泛深入地发展，故乐为此序。

李家治

1997 年 5 月 18 日于上海

前 言

近年来多元统计分析方法已涉及许多研究领域，特别是应用于那些曾经被人们认为与数学无缘的学科，诸如：地质、地震、天气预报、医学以及考古等，取得了许多令人瞩目的成就。计算机的日益普及又为这些学科的研究注入了新的活力，导致了被称之为“数学地质”以及“计算机考古 (Computer Archaeology)”等交叉学科的出现。现有研究资料表明，多元统计分析的应用范围将会进一步扩大，研究问题的深度和广度也将进一步增加。

中国古陶瓷在世界享有盛誉，对人类文明的发展与进步作出了极大的贡献。如何研究并总结这伟大的文明遗存，不但是我们中华民族责无旁贷的职责，而且也引起了其他许多国家科学家的研究兴趣。但是，古陶瓷的研究涉及到许多学科，研究问题非常复杂，如何对所研究的问题进行归纳、总结和提炼则是每一个研究者所面临的问题，而归纳总结的情况又直接影响到研究结果的优劣。因此，本书以中国古陶瓷物理化学及中国古陶瓷科技发展史为讨论的主要对象，向读者展示多元统计分析的数学方法、实用的计算机应用程序 (FORTRAN) 以及应用技巧，试图为古陶瓷及相关学科的科研人员、高校教师与研究生提供参考资料。

本书所涉及的多元统计分析方法及所给出的实用程序除可用于书中叙述的研究范围外，还可用于古陶瓷器型、纹饰以及考古发掘物的分类等方面的研究。由于篇幅有限，恕作者未给出这些方面的例子。

除前言与绪论外，全书共包括 4 个部分内容：(1) 多元统计分析数学方法；(2) 中国古陶瓷物理化学基础与多元统计分析；(3) 中国古陶瓷化学成分数据库；(4) 多元统计分析软件使用方法及程序清单。

此书的主要内容是作者已发表与未发表科研工作的总结，同时也包含我的导师李家治、高力明和游恩溥教授以及其他科研人员的辛勤劳动，特别是其中的聚类分析程序是由高力明教授提供的。另外，还参阅了国内外已发表的有关论文及著作，并承蒙多位同志审阅，提出了许多好的建议及修改意见，国家自然科学基金委员会、中国轻工业出版社在出版等方面给予了大力支持，作者在此一并表示由衷的感谢。

由于作者水平有限，书中错误与不足在所难免，恳请各位读者不吝赐教。

罗宏杰

1997 年 5 月于西北轻工业学院

目 录

概述.....	1
第一章 多元统计分析的数学方法.....	9
第一节 聚类分析与最小支撑树分析.....	9
第二节 对应分析	14
第三节 Bayes 准则多类逐步判别分析	18
第二章 中国古陶瓷物理化学基础与多元统计分析	22
第一节 中国古陶瓷发展过程的研究	22
第二节 景德镇历代陶瓷配方演变过程的研究	60
第三节 浙江地区古陶瓷工艺发展过程的研究	66
第四节 耀州窑古陶瓷工艺演变过程的研究	74
第三章 中国南北方陶器、原始瓷的组成特点及北方出土印纹陶和原始 瓷产地的探讨	83
第一节 南北方陶器及印纹陶的组成特点及北方出土印纹硬陶 产地的推测	83
第二节 北方出土原始瓷产地的研究	86
第四章 中国古瓷中钙系釉类型划分标准及其在瓷釉研究中的应用	93
第一节 草木灰釉式图	93
第二节 不同类型钙系釉划分标准的讨论	94
第三节 钙系釉的划分标准与南北方青、白瓷釉	95
第五章 聚类与判别分析在中国古陶瓷研究中的应用	98
第一节 景德镇地区古瓷	98
第二节 浙江地区古陶瓷.....	105
第三节 河南临汝和耀州窑青瓷.....	111
第四节 同一地区不同时期瓷釉的判别函数.....	113
第五节 不同时期耀州窑青瓷的判别函数.....	116
第六章 中国古陶瓷化学成分数据库简介.....	122
第一节 数据库结构.....	122
第二节 数据库使用说明简介.....	124
第三节 中国古陶瓷化学成分汇总.....	125
第七章 多元统计分析程序清单及使用说明.....	129
第一节 聚类分析与最小支撑树分析.....	129

第二节 对应分析.....	142
第三节 判别分析.....	150
附表 1 中国古陶瓷胎化学成分汇总表	162
附表 2 中国古陶器涂层及瓷釉化学成分汇总表	206

概 述

一、古陶瓷科技发展的研究进展

中国是世界著名的陶瓷生产地，并于汉代至五代（公元前 206 年至公元 960 年）在全世界范围内率先发明了瓷器。从我国第 1 件陶器制品的诞生，到今天种类繁多的陶瓷制品（包括传统陶瓷与新型陶瓷）的出现，经历了长达 1 万余年的历史演变过程。每一件陶瓷产品都是我国劳动人民智慧与文明的结晶，都记录着我国文明与科技的发展过程，它们是我们今天研究历史、文化及陶瓷科技演变史的宝库。

本世纪 20 年代，周仁先生率先在我国开展了利用自然科学方法进行中国古陶瓷研究的工作，为我国古陶瓷的研究和一些名窑陶瓷产品的恢复做出了不可磨灭的贡献。继周仁先生之后，许多科技工作者在古陶瓷的研究领域中都曾或正在进行着大量的工作，取得了许多令人鼓舞的成绩。

传统陶瓷是由天然粘土、长石、石英、瓷石及一些其他辅助天然原料按不同配比配制，并经原料加工、成型、烧成而制得的制品。由于我国幅员辽阔，不但南北地理环境相差较大，使天然原料的组成及物理性能有较大的差异，而且就南方或北方而言，原料的局部地域性也会造成原料的性能有一定的差异，再加之陶瓷胎釉配方、制作工艺（包括原料的细度、烧成制度）等因素的差别，使得我国陶瓷的化学组成及结构非常复杂。这就为对大量陶瓷组成及结构的综合比较与总结带来了相当大的困难。

在古陶瓷研究工作中，我们期望通过研究它们的化学组成及显微结构，而得到古代陶瓷制作的一些信息。从陶瓷工艺及物理化学的角度来看，陶瓷的化学组成是陶瓷制作所采用原料的化学组成及其原料配比的函数；而陶瓷显微结构则不但取决于制作原料的化学组成及原料配比，而且取决于陶瓷的制作工艺。因而，可通过比较不同古陶瓷样品的化学组成及显微结构来讨论古代陶瓷的制作原料、配方及其工艺特点。

传统陶瓷化学组成包含 SiO_2 、 Al_2O_3 、 K_2O 、 Na_2O 、 CaO 、 MgO 、 Fe_2O_3 、 TiO_2 等多种氧化物。若以每一种氧化物为坐标画图时，每一个陶瓷胎或釉的化学组成则对应于多个氧化物所构成的多维空间中的一个点。一般，陶瓷胎或釉化学组成中的氧化物种类远大于 2，因而若不经其他方法处理时，则很难用大家熟悉的二维平面图来比较不同陶瓷胎釉化学组成之间的异同点，这为人们比较不同陶瓷胎釉化学组成增加了相当大的难度。长期以来，人们一直在寻求简单明了的解决方法。Seger 化学式及其演变得来的胎式或釉式的出现，使人们能够利用二维平面图很直观地比较不同陶瓷胎釉化学组成的异同，因而，备受陶瓷工作者的重视。Seger 胎式与釉式的计算方法如下所示：

胎式： $a \text{ R}_2\text{ORO} \cdot b \text{ R}_2\text{O}_3 \cdot c \text{ RO}_2$

釉式： $\text{R}_2\text{ORO} \cdot c \text{ R}_2\text{O}_3 \cdot d \text{ RO}_2$

式中, R_2O 代表一价碱金属氧化物, 包括 K_2O 、 Na_2O 等; RO 代表二价碱土金属氧化物, 包括 CaO 、 MgO 等; R_2O_3 代表三价金属氧化物, 包括 Al_2O_3 、 Fe_2O_3 等; 而 RO_2 代表四价金属氧化物, 包括 SiO_2 、 TiO_2 等。在胎式中通常令中性氧化物 R_2O_3 的系数为 1, 故 a 、 b 分别表示 R_2O 、 RO 及 RO_2 的分子数相对于 R_2O_3 分子数的比值。而在釉式的计算中, 一般都令碱性氧化物 R_2ORO (包括碱金属和碱土金属) 的系数为 1, c 、 d 分别表示 R_2O_3 、 RO_2 的分子数相对于 R_2ORO 分子数的比值。

当以 a 对 b (胎式) 或 c 对 d (釉式) 作图时, 每一个陶瓷胎或釉的化学组成就正对应二维平面上的一个点。因此可利用胎式或釉式图比较不同陶瓷胎或釉在化学组成上的异同点, 这就初步解决了在多维空间中, 不同陶瓷胎或釉化学组成不能比较的问题。

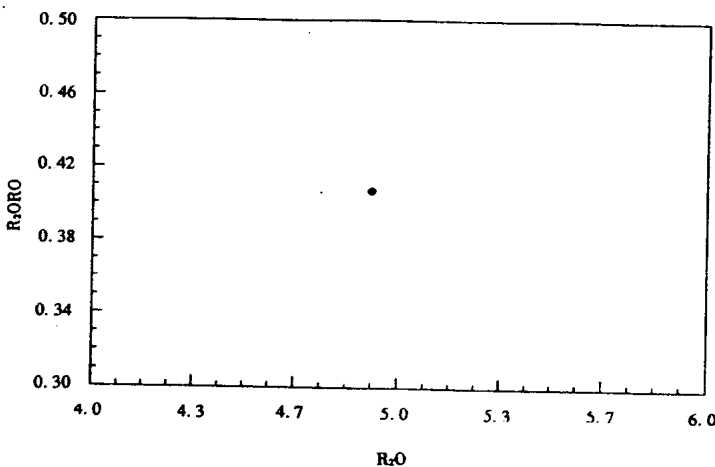


图 1 样品胎式图

No.1 到 No.6 样品的碱性氧化物 (CaO 、 MgO 、 K_2O 、 Na_2O) 含量有较大的差异, 但它们的胎式则完全相同。因而, 用胎式或釉式比较这些样品时 (见图 1), 就看不出这 6 个样品在化学组成上的差异, 容易掩盖同一类氧化物之间的差别。

在比较不同陶瓷胎釉化学组成异同点时, Seger 法确实具有较大的优势, 且也在古陶瓷的研究应用中起到了较大的作用。但是, 这种方法仍然存在一些不足之处。可以看出, 在上述胎、釉式的计算中, 由于将不同的氧化物按其分子式的特点给予了合并, 这也就掩盖了同类氧化物 (即碱性氧化物、中性氧化物及酸性氧化物) 间的差异对比较结果的影响。从表 1 可以看出,

表 1 样品的化学组成及胎式计算结果

序号	化 学 成 分						胎 式	
	SiO_2	Al_2O_3	CaO	MgO	K_2O	Na_2O	R_2ORO	RO_2
No.1	69.5	24.0	0.00	1.2	3.5	1.8	0.407	4.914
No.2	69.5	24.0	1.80	0.98	3.72	0.0	0.407	4.914
No.3	69.5	24.0	2.30	0.73	3.47	0.0	0.407	4.914
No.4	69.5	24.0	2.80	0.48	3.22	0.0	0.407	4.914
No.5	69.5	24.0	3.30	0.22	2.98	0.0	0.407	4.914
No.6	69.5	24.0	3.74	0.00	2.76	0.0	0.407	4.914

注: 本书表格中化学组成和配方均以质量分数表示。

如何用广大研究者所熟悉的二维平面图在不明显损失数据所固有信息的基础上,研究陶瓷样品化学组成间的关系,显得重要而又迫切。国内外在地质、考古等方面的研究表明,多元统计分析方法完全可应用于陶瓷化学组成比较方面的研究。

多元统计分析方法最早应用于心理学等方面的研究,取得了良好的效果,因而备受统计学家及其他科学家的注意。本世纪 50 年代以后,多元统计分析方法被用于地质学的研究领域之中,诸如:利用时间趋势进行地层对比分析;利用因子分析研究沉积环境等,并由此而形成了一门新兴的学科——数学地质。近年来,随着计算机应用的不断普及,多元统计分析方法已应用到许多研究领域,并取得了许多好的研究结果。

电子计算机的出现及其普及进一步开拓了多元统计分析方法的应用范围。许多统计计算只需几分钟,甚至几秒钟就能完成。电子计算机的应用大大推动了多元统计分析方法的发展,同时多元统计分析方法的进一步发展完善又为更多领域内使用计算机以及多元统计分析方法开辟了广阔的道路。当一些较新、较好的数学方法与古老的学科相结合时,就会使这些学科的研究状况发生很大的变化,并由此孕育了许多新的边缘学科,如地质与数学结合为数学地质,考古与计算机及多元统计分析等数学方法结合为计算机考古。

当我们用对应分析方法处理表 1 中所列数据,并将处理结果标绘在图 2 时,6 个样品的组成点并不重合在一起。这些样品的化学组成特点,可通过分析样品点与氧化物组成点的亲疏关系来得到(具体分析方法见计算机辅助实验结果讨论部分)。这正反映了多元统计分析方法在此研究方面的优势。

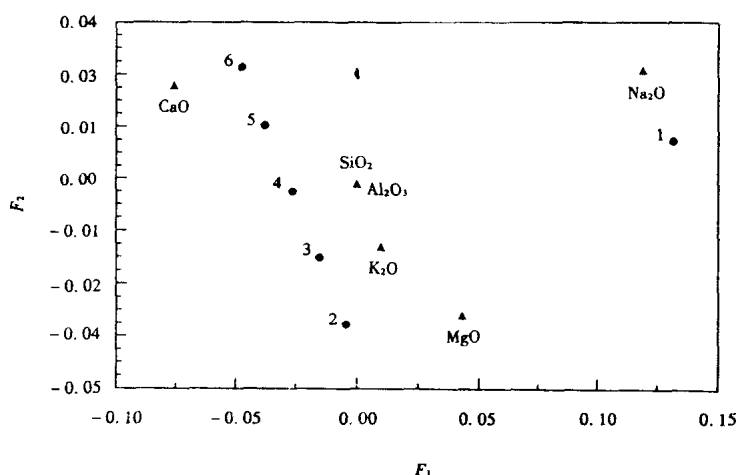


图 2 样品 F_1 、 F_2 因子载荷分布图

二、多元统计分析的计算机辅助实验

多元统计分析方法固然可根据化学组成来区分不同陶瓷样品,但能否通过研究多元统计分析的结果而了解陶瓷样品制作工艺方面的一些信息,则是古陶瓷研究者所关心的问题。对此,可用下面的计算机辅助实验予以说明。

选用 10 种化学组成不同的瓷石原料和 1 种高岭土设计陶瓷胎的化学组成(配方及

化学组成见表 2)，并采用聚类分析及对应分析方法（数学方法见第一章）分别处理表 2 中所列实验样品的化学组成数据，结果分别展示在图 3、图 4 及图 5 中。

表 2 二元配料系统计算机实验配方及其化学组成

序 号	配 方		化 学 组 成										合计
	瓷 石	高岭土	SiO ₂	Al ₂ O ₃	TiO ₂	Fe ₂ O ₃	K ₂ O	Na ₂ O	CaO	MgO	MnO		
1	瓷石 1 * 100		80.33	15.15	0.00	0.45	3.21	0.35	0.36	0.16	0.00	100.01	
2	瓷石 2 100		78.44	15.30	0.00	0.69	2.71	1.30	1.05	0.47	0.04	100.00	
3	瓷石 3 100		78.84	15.50	0.00	0.79	2.87	0.44	1.50	0.00	0.06	100.00	
4	瓷石 4 100		78.05	15.08	0.00	0.61	2.09	3.23	0.54	0.31	0.09	100.00	
5	瓷石 5 100		77.76	14.51	0.00	0.60	1.96	4.26	0.52	0.29	0.09	99.99	
6	瓷石 6 100		79.98	14.53	0.00	0.60	2.83	0.52	1.15	0.31	0.07	99.99	
7	瓷石 7 100		77.66	16.19	0.13	1.18	3.97	0.56	0.03	0.24	0.04	100.00	
8	瓷石 8 100		80.15	14.39	0.07	0.85	3.52	0.19	0.38	0.46	0.00	100.01	
9	瓷石 9 100		76.22	16.30	0.09	0.58	3.92	0.61	1.90	0.36	0.02	100.00	
10	瓷石 10 100		72.86	18.39	0.07	0.69	4.80	0.56	2.20	0.42	0.01	100.00	
11	瓷石 10 40, 瓷石 5 60		74.82	16.84	0.04	0.65	3.66	2.04	1.53	0.37	0.04	99.99	
12	瓷石 1 90	10	77.65	17.68	0.00	0.51	3.17	0.42	0.37	0.17	0.02	99.99	
13	瓷石 2 90	10	75.96	17.81	0.00	0.73	2.72	1.27	0.99	0.45	0.05	99.98	
14	瓷石 3 90	10	76.32	18.00	0.00	0.82	2.86	0.50	1.40	0.03	0.07	100.00	
15	瓷石 4 90	10	75.60	17.62	0.00	0.66	2.16	3.01	0.53	0.30	0.10	99.98	
16	瓷石 5 90	10	75.34	17.11	0.00	0.65	2.05	3.94	0.51	0.29	0.10	99.99	
17	瓷石 6 90	10	77.34	17.13	0.00	0.65	2.83	0.57	1.08	0.31	0.08	99.99	
18	瓷石 7 90	10	75.25	18.62	0.12	1.17	3.85	0.61	0.07	0.24	0.05	99.98	
19	瓷石 8 90	10	77.49	17.00	0.07	0.87	3.45	0.28	0.39	0.44	0.02	100.01	
20	瓷石 9 90	10	73.95	18.72	0.09	0.63	3.81	0.66	1.76	0.35	0.03	100.00	
21	瓷石 10 90	10	70.93	20.60	0.07	0.73	4.60	0.61	2.02	0.40	0.03	99.99	
22	瓷石 10 40, 瓷石 5 50	10	72.89	19.05	0.04	0.70	3.47	2.09	1.35	0.35	0.06	100.00	
23	瓷石 1 75	25	73.64	21.48	0.01	0.61	3.11	0.53	0.38	0.19	0.04	99.99	
24	瓷石 2 75	25	72.23	21.59	0.01	0.80	2.74	1.24	0.90	0.43	0.07	100.01	
25	瓷石 3 75	25	72.53	21.75	0.01	0.87	2.86	0.59	1.24	0.07	0.09	100.01	
26	瓷石 4 75	25	71.93	21.43	0.01	0.74	2.27	2.69	0.52	0.30	0.11	100.00	
27	瓷石 5 75	25	71.72	21.00	0.01	0.73	2.18	3.46	0.50	0.29	0.11	100.00	
28	瓷石 6 75	25	73.39	21.02	0.01	0.73	2.83	0.66	0.98	0.30	0.09	100.01	
29	瓷石 7 75	25	71.64	22.26	0.11	1.16	3.68	0.69	0.13	0.25	0.07	99.99	
30	瓷石 8 75	25	73.51	20.91	0.06	0.91	3.35	0.41	0.40	0.41	0.04	100.00	
31	瓷石 9 75	25	70.56	22.34	0.08	0.71	3.65	0.72	1.54	0.34	0.05	99.99	
32	瓷石 10 75	25	68.04	23.91	0.07	0.79	4.31	0.69	1.76	0.38	0.05	100.00	
33	瓷石 10 40, 瓷石 5 35	25	70.01	22.36	0.03	0.76	3.17	2.17	1.09	0.33	0.08	100.00	
34	瓷石 1 55	45	68.30	26.54	0.02	0.75	3.04	0.67	0.40	0.21	0.07	100.00	
35	瓷石 2 55	45	67.26	26.62	0.02	0.88	2.76	1.19	0.78	0.39	0.09	99.99	
36	瓷石 3 55	45	67.48	26.74	0.02	0.93	2.85	0.72	1.03	0.13	0.10	100.00	
37	瓷石 4 55	45	67.04	26.51	0.02	0.84	2.42	2.26	0.50	0.29	0.12	100.00	
38	瓷石 5 55	45	66.89	26.19	0.02	0.83	2.35	2.82	0.49	0.29	0.12	100.00	
39	瓷石 6 55	45	68.11	26.20	0.02	0.83	2.83	0.77	0.84	0.30	0.11	100.01	
40	瓷石 7 55	45	66.83	27.12	0.09	1.15	3.45	0.79	0.22	0.26	0.09	100.00	
41	瓷石 8 55	45	68.20	26.12	0.06	0.97	3.21	0.58	0.41	0.38	0.07	100.00	
42	瓷石 9 55	45	66.04	27.18	0.07	0.82	3.43	0.82	1.25	0.32	0.08	100.01	
43	瓷石 10 55	45	64.19	28.33	0.06	0.88	3.91	0.79	1.41	0.36	0.08	100.01	
44	瓷石 10 40, 瓷石 5 15	45	66.15	26.77	0.03	0.84	2.77	2.27	0.74	0.31	0.11	99.99	
45		100	53.60	40.47	0.04	1.11	2.82	1.07	0.45	0.28	0.16	100.00	

* 瓷石 1, 瓷石 2..... 分别代表第 1, 2..... 种瓷石。

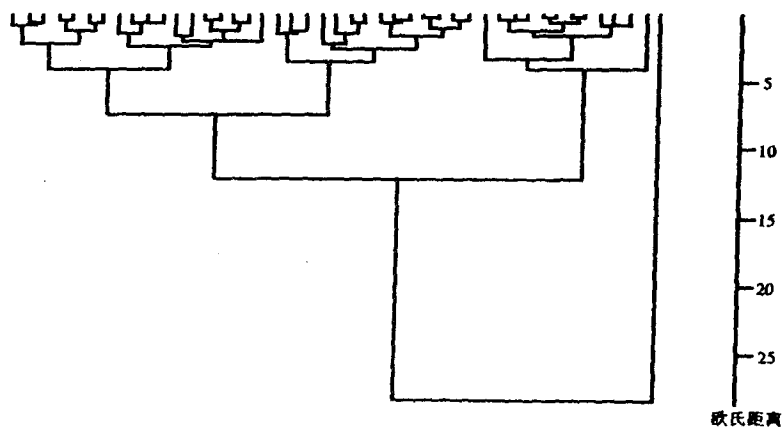


图3 计算机辅助实验样品 Q-模式聚类谱系图

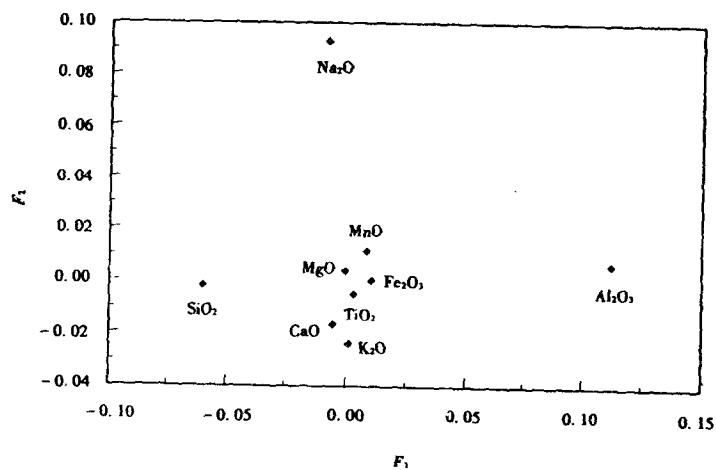


图4 计算机辅助实验样品 R-模式因子载荷分布图

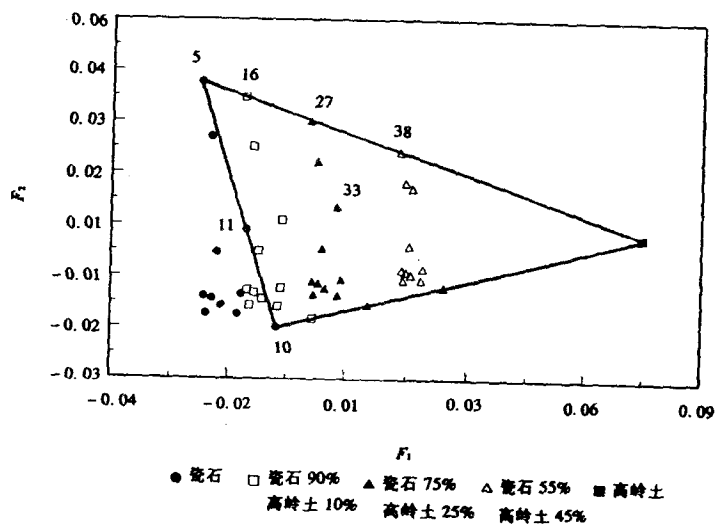


图5 计算机辅助实验样品 Q-模式因子载荷分布图

当以阈值 $F=4.0$ 对图 3 中的谱系图分隔后, 所有样品被分为 5 类。若将表 2 中序号为 1~11、12~22、23~33、34~44 及 45 样品的配方分别称为配方 1、配方 2……配方 5 时, 每一类样品所对应的配方信息及聚类结果见表 3。

表 3 聚类分析结果统计表

聚类类别	序号	配 方		聚类类别中样品数 (占同类配方样品总数百分数)				
		瓷 石	高岭土	配方 1	配方 2	配方 3	配方 4	配方 5
I 类	1	瓷石 1 100		7 (63.64%)				
	6	瓷石 6 100						
	8	瓷石 8 100						
	2	瓷石 2 100						
	3	瓷石 3 100						
	4	瓷石 4 100						
II 类	5	瓷石 5 100						
	7	瓷石 7 100		3 (27.27%)	8 (72.72%)			
	12	瓷石 1 90	10					
	17	瓷石 6 90	10					
	19	瓷石 8 90	10					
	9	瓷石 9 100						
	11	瓷石 10 40, 瓷石 5 60						
	13	瓷石 2 90	10					
	14	瓷石 3 90	10					
	15	瓷石 4 90	10					
	16	瓷石 5 90	10					
III 类	18	瓷石 7 90	10					
	10	瓷石 10 100		1 (9.09%)	3 (27.28%)	10 (90.91%)		
	20	瓷石 9 90	10					
	22	瓷石 10 40, 瓷石 5 50	10					
	21	瓷石 10 90	10					
	29	瓷石 7 75	25					
	31	瓷石 9 75	25					
	33	瓷石 10 40, 瓷石 5 35	25					
	23	瓷石 1 75	25					
	28	瓷石 6 75	25					
	30	瓷石 8 75	25					
	24	瓷石 2 75	25					
	25	瓷石 3 75	25					
	26	瓷石 4 75	25					
IV 类	27	瓷石 5 75	25					
	32	瓷石 10 75	25			1 (9.09%)	11 (100.0%)	
	34	瓷石 1 55	45					
	39	瓷石 6 55	45					
	41	瓷石 8 55	45					
	35	瓷石 2 55	45					
	36	瓷石 3 55	45					
	37	瓷石 4 55	45					
	38	瓷石 5 55	45					
	40	瓷石 7 55	45					
	42	瓷石 9 55	45					
	44	瓷石 10 40, 瓷石 5 15	45					
V 类	43	瓷石 10 55	45					
	45		100					1 (100%)

注: 表中自上而下的样品序号与聚类谱系图 3 自上而下的样品序号相同。

由表 3 可见, 计算机辅助实验数据的聚类分析结果不但能反映样品的聚类特征, 而且也能反映样品的配方特征。

图 4 和图 5 分别是 R-模式和 Q-模式对应分析 F_1 和 F_2 因子载荷分布图。 F_1 、 F_2 因子方差贡献分别为 89% 和 10%。因子方差贡献愈大表明因子所反映的研究信息愈多。通过分析样品的因子载荷分布图, 可以了解样品、变量以及样品与变量之间的关系, 从而进一步了解陶瓷形成的物理化学内涵及不同类陶瓷样品的制作工艺特点等。

样品间关系 所有样品点都位于配制该样品所使用的两种原料的连线上 (如 16、27、38 号样品都是由瓷石 5 和高岭土配制而得的), 或者位于配制该样品所使用 3 种原料的三角形内 (如 22、33、44 号样品是由瓷石 10、瓷石 5 和高岭土所配制的)。由样品在连线或三角形内的位置可准确推得样品的实际配方。

变量间关系 对于具有较大载荷的变量来说, 相互距离的远近是它们关系密切程度的反映, 也是样品配方及配方中原料组成特点等信息的综合反映。 SiO_2 和 Al_2O_3 (在 F_1 因子中) 以及 Na_2O 和 K_2O (在 F_2 因子中) 组成点的距离都较远, 且都呈负相关关系 (见表 4), 这表明样品 SiO_2 含量的增加, 会导致其 Al_2O_3 含量的降低, 同时 K_2O 含量的增加也会使其 Na_2O 含量降低。 SiO_2 和 Al_2O_3 的负相关关系是样品配方特征信息的反映, 表示配方中的富 Al_2O_3 原料 (高岭土) 的用量增加时会导致富 SiO_2 原料 (瓷石) 用量的减少。同时, Na_2O 和 K_2O 与 SiO_2 组成点的距离较之与 Al_2O_3 组成点间的距离为近, 这表明 Na_2O 和 K_2O 主要是由富 SiO_2 原料 (瓷石) 所提供的, 样品含 K_2O 、 Na_2O 量的多少则是瓷石类型变化的反映。综上所述可见, 如果样品的组成点位于靠近 SiO_2 和 K_2O 一侧, 不但表明样品配方中瓷石的含量较高, 而且也进一步表明了瓷石的类型为富 K_2O 瓷石。

表 4 R-模式因子载荷表

变 量	F_1	F_2
SiO_2	-0.060314	-0.005577
Al_2O_3	0.112218	0.003263
TiO_2	0.003103	-0.008679
Fe_2O_3	0.010179	-0.003232
K_2O	0.001244	-0.028109
Na_2O	-0.008548	0.092305
CaO	-0.005527	-0.020609
MgO	-0.000344	0.000629
MnO	0.008039	0.008361
方差累计值	52.58%	84.20%

样品和变量间的关系 (将图 5 迭加于图 4 后可讨论样品与变量间的关系) 当样品和载荷较大的变量 (或氧化物) 相距较近时, 表明该样品富含该变量 (或氧化物)。所有瓷石的组成点都靠近 SiO_2 而远离 Al_2O_3 的组成点, 这表明瓷石富含 SiO_2 而贫 Al_2O_3 。

高岭土的组成点较其他所有样品都接近 Al_2O_3 的组成点, 表明高岭土的 Al_2O_3 含量在所研究的样品中是最高的。另外, 分析样品组成点与 Na_2O 、 K_2O 的距离关系可推测样品中这些氧化物的含量。

从上述分析可以看出, 聚类分析与对应分析方法不但能对研究的样品按化学组成进行分类, 而且可进一步确定样品配方, 并推测样品配方中所使用原料的某些特性, 完全适应古陶瓷研究的需要。

参 考 文 献

1. 李家治. 我国瓷器出现时期的研究. 硅酸盐学报, 1978, 6 (3): 190~198
2. 於崇文. 数学地质的方法与应用. 北京: 冶金出版社, 1980
3. 王学仁. 地质数据的多变量分析. 北京: 科学出版社, 1986, 306~316
4. 罗宏杰. 多元统计分析方法在景德镇古瓷配方、鉴别研究中的应用与中国古陶瓷化学成分数据库的建立 (硕士论文). 咸阳: 西北轻工业学院, 1988
5. Dons Rice. 新考古学. 交流, 1986, 3
6. J.D. Wilcock. The History and Development of Computer Archaeology. (来华讲稿)
7. Garman Harbottle. Provenience Studies Using Neutron Activation: The Role of Standardization. Archaeology Ceramics, 1982
8. Gao Liming, Luo Hongjie, John Wilcock. The analysis of ancient Chinese pottery and porcelain shapes. In: computer applications and quantitative methods in archaeology, 1989. 363

第一章 多元统计分析的数学方法

多元统计分析所包括的内容十分广泛,本章只叙述作者近几年研究所采用的聚类分析、因子分析、对应分析以及判别分析等多元统计分析方法的数学方法。另外,为了大家的使用方便,书中给出了这些多元统计分析方法的 FORTRAN 应用程序清单及应用例证。阅读多元统计分析的数学方法时,需要读者掌握线性代数、概率论以及数理统计的基本知识。

在叙述多元统计分析方法及其应用例证时,我们以样品代替所研究的古陶瓷碎片,而把陶瓷研究的指标(如陶瓷的化学组成及物理性能等)通称为变量。

第一节 聚类分析与最小支撑树分析

聚类分析属于一种多元分类分析方法,通过聚类分析可以将样品及变量归并为若干类别,使得每一类内的所有个体之间具有较密切的关系,而类别之间的关系则相对比较疏远。

依据分类目的的差异,聚类分析可分为两类,一类是对样品进行分类,一般称为 Q-模式分析;另一类对变量进行分析,称为 R-模式分析。由于篇幅有限,此处仅介绍 Q-模式聚类分析的步骤及数学方法。

一、系统聚类分析

(一) 聚类分析步骤

1. 各样品自成一类(这时有 n 类)。
2. 计算各个样品之间的距离(计算方法见本节第二部分)。将最近的两个样品并成一类。
3. 计算新类与其余各类的距离(类之间距离的定义方法及计算见本节第三部分),再将距离最近的两类合并。这时如果类的个数仍大于 1,则继续重复上述步骤,直到所有样品归为一类时停止计算。

(二) 样品间距离的定义及计算

设有 n 个样品 $x_1, x_2, \dots, x_n$, 各测得 m 个变量 $X_1, X_2, \dots, X_m$, 用 X_{ij} 表示第 i 个样品的第 j 个变量,通常数据列成表 1-1 的形式。

如果将样品 $x_1, x_2, \dots, x_n$ 看成是 m 维空间的 n 个点,从几何的角度来看,可以定义这些点之间的距离。样品 x_i 和 x_j 的欧氏(Euclidean)距离可用下式表示(本书所涉及的距离均采用欧氏距离,距离也有其他的定义方法,可参见聚类分析的相关资料):

表 1-1

数 据 阵

变 量 样 品	X_1	X_2	...	X_j	...	X_m
x_1	x_{11}	x_{12}	...	x_{1j}	...	x_{1m}
x_2	x_{21}	x_{22}	...	x_{2j}	...	x_{2m}
$\vdots$	$\vdots$	$\vdots$	...		...	$\vdots$
x_i	x_{i1}	x_{i2}	...	x_{ij}	...	x_{im}
$\vdots$	$\vdots$	$\vdots$	...		...	$\vdots$
x_n	x_{n1}	x_{n2}	...	x_{nj}	...	x_{nm}
均 值	$\bar{x}_1$	$\bar{x}_2$	...	$\bar{x}_j$	...	$\bar{x}_m$
极 差	R_1	R_2	...	R_j	...	R_m
标准差	S_1	S_2	...	S_j	...	S_m

$$d_{ij} = \left[\sum_{k=1}^m (x_{ik} - x_{jk})^2 \right]^{1/2}$$

在上式中变量的量纲对距离的计算有一定的影响, 为了克服这个缺点常将数据按以下两种方法进行标准化。

1. 标准差标准化

$$x'_{ij} = \frac{x_{ij} - \bar{x}_j}{S_j}$$

$$i=1, 2, 3, \dots, n; j=1, 2, 3, \dots, m$$

式中 $\bar{x}_j$ 为 n 个样品第 j 个变量的子样均值, S_j 为 n 个样品第 j 个变量的子样标准差, 它们可通过下式予以计算。

$$\bar{x}_j = \frac{1}{n} \sum_{i=1}^n x_{ij}$$

$$S_j^2 = \frac{1}{n-1} \sum_{i=1}^n (x_{ij} - \bar{x}_j)^2, \quad S_j = \sqrt{S_j^2}$$

这时每个变量的子样标准差都化为 1, 标准化后的数据就与变量的量纲没有关系了。有时称经标准差标准化后的距离为马氏 (Mahalanobis) 距离。

2. 极差标准化

$$x'_{ij} = \frac{x_{ij} - \bar{x}_j}{R_j}$$

$$i=1, 2, 3, \dots, n; j=1, 2, 3, \dots, m$$

式中 R_j 为 n 个样品第 j 个变量的子样极差。

$$R_j = \frac{\max}{1 \leq i \leq n} \{x_{ij}\} - \frac{\min}{1 \leq i \leq n} \{x_{ij}\}$$