1

· 偏相关分析C语言源程序
· Bayes判别分析C语言源程序
· 方差最大正交因子解C语言源程序
· 探索性因子分析算例
· Schmidt正交化QR分解法求解典型相关系数、典型相关变量
· Jacobi法求解实对称矩阵全部特征值及对应特征向量C语言源程序

3

· 求一般矩阵全部特征值的Schmidt正交化QR分解法
· 已知特征值求解对应特征向量的直接法
· Schmidt正交化QR分解法求解线性方程组

Multivariate Statistical Analysis

多元统计分析

写于大数据、云计算时代

李庆来 编著

上海交通大学出版社
SHANGHAI JIAO TONG UNIVERSITY PRESS

Multivariate Statistical Analysis

多元统计分析

写于大数据、云计算时代

李庆来◎编著

- 偏相关分析C语言源程序
- Bayes判别分析C语言源程序
- 方差最大正交因子解C语言源程序
- 探索性因子分析算例
- Schmidt正交化QR分解法求解典型相关系数、典型相关变量
- Jacobi法求解实对称矩阵全部特征值及对应特征向量C语言源程序
- 求一般矩阵全部特征值的Schmidt正交化QR分解法
- 已知特征值求解对应特征向量的直接法
- Schmidt正交化QR分解法求解线性方程组

上海交通大學出版社
SHANGHAI JIAO TONG UNIVERSITY PRESS

内容提要

本书全面系统地揭示了多元统计分析的数学原理、方法和应用，主要包括回归分析、偏相关分析、含定性变量的回归分析、逐步回归分析、多类判别分析、逐步判别分析、系统聚类法，主成分分析、方差最大正交因子解、探索性因子分析、典型相关分析等，并提供了一般文献不写的Windows环境下C语言程序设计所有源代码，而这些往往是真正想解决实际问题的人所必需的。

本书可作为普通高等院校高年级本科生、研究生的教材及参考用书，也可供统计工作者、科技人员使用。

图书在版编目(CIP)数据

多元统计分析 / 李庆来编著. —上海：上海交通大学出版社，2015
ISBN 978-7-313-13761-6

Ⅰ. ①多… Ⅱ. ①李… Ⅲ. ①多元分析—统计分析
Ⅳ. ①O212.4

中国版本图书馆CIP数据核字(2015)第217056号

多元统计分析
——写于大数据、云计算时代

编　　著：李庆来
出版发行：上海交通大学出版社　　地　　址：上海市番禺路951号
邮政编码：200030　　电　　话：021-64071208
出 版 人：韩建民
印　　制：上海天地海设计印刷有限公司　　经　　销：全国新华书店
开　　本：787 mm×1092 mm　1/16　　印　　张：23.25
字　　数：556千字
版　　次：2015年9月第1版　　印　　次：2015年9月第1次印刷
书　　号：ISBN 978-7-313-13761-6/O
定　　价：76.70元

目录

CONTENTS

第1章 绪 论

多元统计分析(Multivariate Statistical Analysis)是数理统计学中近几十年来迅速发展的一个分支,大量的自然科学、工程技术和社会科学的实践都已证实了多元统计分析方法是一种很有用的数据处理方法[3~5]。具体地讲,多元统计分析一般是通过试验或观测等途径获得数据,从概率论和数理统计定义的随机变量的数字特征及统计量出发,建立多个随机变量的统计数学模型,运用高等数学、线性代数、矩阵论的知识求解,借助计算机数值计算方法来近似计算,最终用来研究和解释随机变量之间关系的一门综合性的科学。

多元统计分析的历史可以追溯到19世纪[30],但直到20世纪50年代计算机出现以后才有了广泛的应用和发展[31]。今天,多元统计分析在地质、气象、生物、医学、图像处理、经济分析、工程技术等许多领域都取得了成功的应用,这促进了多元统计分析理论的发展,多元统计分析方法也逐渐成为人们认识世界的一种重要方法。

1.1 多元统计分析方法概述

多元统计分析是数理统计学的一个分支,研究的对象是随机现象,与经典数学方法研究的对象不同,后者研究的是确定性现象。

1.1.1 确定性现象和随机现象

客观世界中发生的现象有两种:确定性现象(如水在0℃时会结冰)和随机现象(如掷一枚硬币有可能出现正面,也可能出现反面)。经典数学方法一般是用来研究确定性现象的,对随机现象无能为力[66]。越来越多的实践说明,概率统计方法与经典数学方法是相辅相成的。

概率论与数理统计是研究和揭示随机现象统计规律性的一门数学学科,多元统计分析则是数理统计学的一个分支。有些多元统计分析文献中的方差分析和回归分析内容,在大学概率论与数理统计教材中均有介绍,因为其中的回归分析一般是一元线性回归分析,所以本书将多元线性回归分析作为多元统计分析中的一种方法来介绍。

1.1.2 多元统计分析方法是一种数据处理方法

多元统计分析方法实际上是一种数据处理方法,这些数据有可能来自试验,也可能来自

观测(包括天文观测、工程测试、社会活动的数据采集等)。数理统计是运用概率论的基本知识,研究如何合理地获得这些数据,是"定量分析"的关键学科[5]。一般来讲,多元统计分析的方法要基于数理统计的理论、方法,只有这样获得的数据,对这些数据的处理结果才有可能给出合理的解释。数理统计的内容一般包括随机变量的统计量及其分布、参数估计、假设检验和方差分析等,这些不作为本书介绍的内容。

多元统计分析方法一般都是处理二维数据的,如表 1.1 所示。

表 1.1 多元数据的统计表

编 号	矩 阵 $\boldsymbol{X}$		
	x_1	x_2	x_3
1	0.2	5.2	0.500
2	1.6	6.1	0.545
3	0.4	9.4	0.762
4	0.2	17.0	1.053
5	0.1	9.7	1.301

在多元统计分析中,类似于表 1.1 的数据可用矩阵表示,如 m 个变量、n 个样品的表格用矩阵表示为

$$\boldsymbol{X}=\begin{pmatrix} x_{11} & x_{12} & \cdots & x_{1m} \\ x_{21} & x_{22} & \cdots & x_{2m} \\ \vdots & \vdots & & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{nm} \end{pmatrix}_{n\times m}=(x_{ij})_{n\times m}(i=1,2,\cdots,n;\ j=1,2,\cdots,m)$$

该矩阵中的元素由 i 行样品、j 列变量组成,在计算机中可用二维数组来表示。

另外,多元统计分析方法所处理的数据都是基于 Euclid 空间(欧式空间)的,并不涉及 Unitary 空间(酉空间)。因此,多元统计分析所涉及的矩阵、随机变量的数字特征和各类统计量是属于 Euclid 空间的,一般都是实数的运算,不会出现复数运算的情况。对于第 9 章所涉及的复数特征值及对应特征向量问题,仅作为参考阅读的内容。

1.1.3 多元统计分析方法的分类

常用的多元统计分析方法有:回归分析、判别分析、聚类分析、主成分分析、对应分析、因子分析和典型相关分析等。本书涉及的多元统计分析方法如表 1.2 所示。

表 1.2 多元统计分析方法

编 号	类 型	章 节	多元统计分析方法	程 序
Ⅰ	线性模型	第 2 章	回归分析、偏相关分析、含定性变量的回归分析、逐步回归分析	有
Ⅱ	分类和分组问题	第 3 章	多类判别分析、逐步判别分析	有
		第 4 章	系统聚类法	—
		第 5 章	主成分分析	有

续 表

编 号	类 型	章 节	多元统计分析方法	程 序
Ⅲ	协方差结构分析	第6章	方差最大正交因子解、探索性因子分析	有
		第7章	典型相关分析(根据相关系数、协方差矩阵计算)	有

注：多元统计分析涉及的线性方程组和矩阵特征值等问题求解的数值算法见第8章和第9章。

1.2 高等数学的相关理论和方法

多元统计分析虽然是数理统计学的一个分支，但它离不开高等数学的基础理论和方法，如随机变量的数字特征和统计分布就是基于高等数学的严格定义；同时，多元统计分析数学模型的求解常常涉及极值问题，需要借助高等数学的求解方法，如最小二乘法(见第2章多元线性回归分析)和Lagrange(拉格朗日)乘数法等。

本节简要回顾多元函数条件极值问题的一种常用求解方法：Lagrange乘数法，因为它是主成分分析、因子分析和典型相关分析极值问题的求解方法。

一般的，求 p 个变量 $x_1, x_2, \cdots, x_p$ 的函数

$$F = f(x_1, x_2, \cdots, x_p) \tag{1.1}$$

受 q 个条件

$$g_i(x_1, x_2, \cdots, x_p) = 0 \ (i = 1, 2, \cdots, q, q < p) \tag{1.2}$$

约束的极值，这类问题称为条件极值问题。其中 $f(x_1, x_2, \cdots, x_p)$ 称为目标函数，$g_i(x_1, x_2, \cdots, x_p) = 0$ 称为约束条件(或附加条件)。

1.2.1 二元函数条件极值问题

对于简单的二元函数条件极值问题，可以通过约束条件解出其中一个变量，然后代入目标函数，根据一元函数极值的求解法则求解。例如：长4 m的绳子围成长为 x、宽为 y 的矩形，求矩形最大面积是多少的问题(根据问题性质可知无极小值)。

$$\begin{cases} F = xy \to \max \\ G = g(x, y) = x + y - 2 = 0 \end{cases}$$

从 G 中解出一个变量 $y = 2 - x$，代入 F 得到 $F = -x^2 + 2x$。根据一元函数极值法则：

$$\frac{\mathrm{d}F}{\mathrm{d}x} = -2x + 2 = 0$$

可得 $x = 1$，$y = 2 - x = 1$，最大面积是 $1\ \mathrm{m}^2$。

1.2.2 Lagrange乘数法

然而，从约束条件中解出一个变量往往是很困难，甚至是不可能的[65]，即使解出了 y，y 也未必是 x 的显式。为了避免上述困难，法国数学家Joseph-Louis Lagrange(约瑟夫·拉

格朗日，1736—1813 年)提出了一种不用解出 y 的方法，后来被称为 Lagrange 乘数法。

仍然用简单的二元函数条件极值问题

$$\begin{cases} F = f(x, y) \to \max \text{ 或 } \min \\ G = g(x, y) = 0 \end{cases}$$

为例，从 G 中解出一个隐函数 $y = y(x)$，代入 F 有 $F = f(x, y(x))$，并对 x 求导得

$$\frac{\mathrm{d}F}{\mathrm{d}x} = \frac{\partial f}{\partial x} + \frac{\partial f}{\partial y} \cdot \frac{\mathrm{d}y}{\mathrm{d}x}$$

在极值点处有

$$\frac{\partial f}{\partial x} + \frac{\partial f}{\partial y} \cdot \frac{\mathrm{d}y}{\mathrm{d}x} = 0 \tag{1.3}$$

将约束条件 $g(x, y) = 0$ 对 x 求导得

$$\frac{\partial g}{\partial x} + \frac{\partial g}{\partial y} \cdot \frac{\mathrm{d}y}{\mathrm{d}x} = 0 \tag{1.4}$$

从式(1.3)和式(1.4)中消去 $\frac{\mathrm{d}y}{\mathrm{d}x}$ 得

$$\frac{\frac{\partial f}{\partial x}}{\frac{\partial g}{\partial x}} = \frac{\frac{\partial f}{\partial y}}{\frac{\partial g}{\partial y}}$$

设此比值为 $-\lambda$，结合约束条件，可得联立方程组：

$$\begin{cases} \dfrac{\partial f}{\partial x} + \lambda \dfrac{\partial g}{\partial x} = 0 \\ \dfrac{\partial f}{\partial x} + \lambda \dfrac{\partial g}{\partial y} = 0 \\ g(x, y) = 0 \end{cases} \tag{1.5}$$

从式(1.5)解出 x，y 和 λ，就得到驻点的坐标(实际求解常常根据问题的性质判断极值是否存在)。

以上二元函数的条件极值求解过程，Lagrange 构造了一个辅助函数：

$$L(x, y) = F + \lambda G = f(x, y) + \lambda g(x, y) \tag{1.6}$$

式中 λ 是待定常数。求 $L(x, y)$ 无条件极值的必要条件，得

$$\begin{cases} \dfrac{\partial L}{\partial x} = \dfrac{\partial f}{\partial x} + \lambda \dfrac{\partial g}{\partial x} = 0 \\ \dfrac{\partial L}{\partial y} = \dfrac{\partial f}{\partial y} + \lambda \dfrac{\partial g}{\partial y} = 0 \end{cases}$$

加上约束条件 $G = g(x, y) = 0$ 就是式(1.5)了。

Lagrange 乘数法可以推广到一般的情形，首先构造形如式(1.6)条件极值问题的辅助

函数：

$$L(x_1, x_2, \cdots, x_p) = f(x_1, x_2, \cdots, x_p) + \sum_{i=1}^{q} \lambda_i g_i(x_1, x_2, \cdots, x_p) \quad (1.7)$$

式中 $\lambda_1, \lambda_2, \cdots, \lambda_q$ 为待定常数，然后求 $L(x_1, x_2, \cdots, x_p)$ 无条件极值的必要条件，结合约束条件，可得联立方程组：

$$\begin{cases} \dfrac{\partial L}{\partial x_j} = \dfrac{\partial f}{\partial x_j} + \sum\limits_{i=1}^{q} \lambda_i \dfrac{\partial g_i}{\partial x_j} = 0 \\ g_i(x_1, x_2, \cdots, x_p) = 0 \end{cases} \quad (i = 1, 2, \cdots, q;\ j = 1, 2, \cdots, p,\ q < p) \quad (1.8)$$

从式(1.8)第一个方程组（p 个极值条件方程）和第二个方程组（q 个约束条件方程）就可以解出 $q+p$ 个未知数 $\lambda_1, \lambda_2, \cdots, \lambda_q, x_1, x_2, \cdots, x_p$，得到可能的极值点坐标。

Lagrange 乘数法的另一个优点是并没有从约束条件中消去某些变量，因此变量 $x_1, x_2, \cdots, x_p$ 具有对称的形式[65]。

下面根据 Lagrange 乘数法，重新求解 1.2.1 节中矩形最大面积问题。

1) 构造辅助函数

根据式(1.6)或式(1.7)构造 Lagrange 辅助函数：

$$L(x, y) = xy + \lambda(x + y - 2)$$

2) 联立方程组

根据式(1.8)得联立方程组：

$$\begin{cases} \dfrac{\partial L}{\partial x} = y + \lambda = 0 \\ \dfrac{\partial L}{\partial y} = x + \lambda = 0 \\ x + y - 2 = 0 \end{cases}$$

从前两个极值条件方程解得 $y = x$，将其代入第 3 个方程 $x + y - 2 = 0$，容易解得：$x = 1$，$y = 1$。

多元统计分析中，主成分分析、因子分析和典型相关分析涉及的条件极值问题的求解使用的就是 Lagrange 乘数法，但求无条件极值的必要条件是对向量的求导，需要借助 1.3 节介绍的矩阵微商公式。

1.3 线性代数和矩阵论知识回顾

多元统计分析依赖于线性代数和矩阵论的理论和方法，主要用于理论推导和科学计算。这里简要回顾一下线性代数和矩阵论的相关知识，更为详细的知识请参见书后所列参考文献。

1.3.1 矩阵和行列式

矩阵和行列式是最容易混淆的两个概念，它们均有严格的数学定义，这里从略。简单

地说，前者就是二维数据，而后者实际就是一个数，但矩阵和行列式还是有很多联系的。本书用“〔 〕”来表示矩阵，用“| |”来表示行列式。例如，2 阶矩阵 $\boldsymbol{A}$ 及其行列式的表示方法为

1）矩阵的表示方法

$$\boldsymbol{A}=\begin{pmatrix}a_{11} & a_{12}\\ a_{21} & a_{22}\end{pmatrix}_{2\times 2}$$

2）矩阵的行列式的表示方法

$$\det(\boldsymbol{A})=\begin{vmatrix}a_{11} & a_{12}\\ a_{21} & a_{22}\end{vmatrix}=a_{11}\,a_{22}-a_{12}\,a_{21}$$

与行列式的表示方法容易混淆的还有变量的绝对值、向量的范数和复数的模。这里，绝对值用“| |”来表示，例如：$|a|=|-3.0|=3.0$；向量的范数用“‖ ‖”来表示，例如向量 $\boldsymbol{e}=(1.0,\ 1.0,\ 0.0)'$ 的范数为

$$\|\boldsymbol{e}\|=\sqrt{1.0^2+1.0^2+0.0^2}=\sqrt{2}$$

而复数的模用“| |”表示，例如复数 1.0＋2.0i 的模：

$$|1.0+2.0\mathrm{i}|=\sqrt{1.0^2+2.0^2}=\sqrt{5.0}$$

有关矩阵和行列式区别和行列式的计算方法详见 8.3 节。

1.3.2 矩阵微商常用公式

1）矩阵对向量的微商定义

场论中梯度的定义可以理解为数量函数 $f(x,\ y,\ z)$ 对向量 $(x,\ y,\ z)$ 的导数：

$$\mathbf{grad}f=\nabla f=\left(\frac{\partial f}{\partial x},\ \frac{\partial f}{\partial y},\ \frac{\partial f}{\partial z}\right)$$

可将这一概念推广到矩阵对向量的微商情形[57]。

设 $x=(x_1,\ x_2,\ \cdots,\ x_m)'$，$y=y(x)=y(x_1,\ x_2,\ \cdots,\ x_m)$ 是以向量 x 为自变量数量函数，即为 m 元函数，则规定 y 对 x 的导数定义为

$$\frac{\mathrm{d}y}{\mathrm{d}x}=\begin{pmatrix}\dfrac{\partial y}{\partial x_1}\\ \dfrac{\partial y}{\partial x_2}\\ \vdots\\ \dfrac{\partial y}{\partial x_m}\end{pmatrix}_{m\times 1}\tag{1.9}$$

当 n 个数量函数组成的矩阵 $\boldsymbol{Y}=y_i(x)=y_i(x_1,\ x_2,\ \cdots,\ x_m)$，$i=1,\ 2,\ \cdots,\ n$ 对 x 的导数定义为

$$\frac{\mathrm{d}\boldsymbol{Y}'}{\mathrm{d}x}=\frac{\partial(y_1,\ y_2,\ \cdots,\ y_n)}{\partial x}=\begin{pmatrix}\dfrac{\partial y_1}{\partial x_1} & \dfrac{\partial y_2}{\partial x_1} & \cdots & \dfrac{\partial y_n}{\partial x_1}\\ \dfrac{\partial y_1}{\partial x_2} & \dfrac{\partial y_2}{\partial x_2} & \cdots & \dfrac{\partial y_n}{\partial x_2}\\ \vdots & \vdots & & \vdots\\ \dfrac{\partial y_1}{\partial x_m} & \dfrac{\partial y_2}{\partial x_m} & \cdots & \dfrac{\partial y_n}{\partial x_m}\end{pmatrix}_{m\times n} \tag{1.10}$$

式中，

$$\boldsymbol{Y}=\begin{pmatrix}y_1\\ y_2\\ \vdots\\ y_n\end{pmatrix}_{n\times 1}=\begin{pmatrix}y_1(x)\\ y_2(x)\\ \vdots\\ y_n(x)\end{pmatrix}_{n\times 1}=\begin{pmatrix}y_1(x_1,\ x_2,\ \cdots,\ x_m)\\ y_2(x_1,\ x_2,\ \cdots,\ x_m)\\ \vdots\\ y_n(x_1,\ x_2,\ \cdots,\ x_m)\end{pmatrix}_{n\times 1} \tag{1.11}$$

根据上述定义有

$$\frac{\mathrm{d}x'}{\mathrm{d}x}=\boldsymbol{I} \tag{1.12}$$

式中，$\boldsymbol{I}$ 为单位矩阵。

另外，设矩阵 $\boldsymbol{A}=(a_{ij})_{n\times n}$ 为常量矩阵，$x=(x_1,\ x_2,\ \cdots,\ x_m)'$，对于数量函数 $y(x)=x'\boldsymbol{A}x$ 对 x 的导数为

$$\frac{\mathrm{d}y}{\mathrm{d}x}=\frac{\partial x'}{\partial x}\boldsymbol{A}x+\frac{\partial x'}{\partial x}\boldsymbol{A}'x=(\boldsymbol{A}+\boldsymbol{A}')x \tag{1.13}$$

当 $\boldsymbol{A}$ 为对称矩阵时有

$$\frac{\mathrm{d}y}{\mathrm{d}x}=\frac{\partial(x'\boldsymbol{A}x)}{\partial x}=2\boldsymbol{A}x \tag{1.14}$$

2）矩阵微商常用公式

当 x，y 为适当的向量，$\boldsymbol{A}$ 为 n 阶矩阵，还有下面的 3 个常用公式[30]。

$$\begin{cases}\dfrac{\partial \boldsymbol{A}x}{\partial x}=\dfrac{\partial x'}{\partial x}\boldsymbol{A}'=\boldsymbol{A}'\\ \dfrac{\partial x'\boldsymbol{A}y}{\partial x}=\dfrac{\partial x'}{\partial x}\boldsymbol{A}y=\boldsymbol{A}y\\ \dfrac{\partial y'\boldsymbol{A}x}{\partial x}=\dfrac{\partial x'}{\partial x}\boldsymbol{A}'y=\boldsymbol{A}'y\end{cases} \tag{1.15}$$

更多有关矩阵的微分和积分公式，详见参考文献[13,57]。

1.4 概率论和数理统计简介

这里仅对均值向量和协方差矩阵作一补充说明，因为主成分分析、因子分析和典型相关分析涉及这方面的知识，在研究协方差矩阵的结构时需要这些公式。

1.4.1 概率论和数理统计研究的主要内容

概率论和数理统计是多元统计分析的基础，它研究的内容包括概率论、数理统计和随机过程等内容。多元统计分析常常涉及随机变量的数字特征（如数学期望、方差、协方差和简单相关系数等）和一些统计分布（正态分布、F 分布、Wilks 统计量、χ^2 分布等），这在概率论和数理统计教材中已有严格定义。实际上，多元统计分析没有偏离这些基本的定义。

1.4.2 随机向量线性组合的协方差阵

考虑 m 个随机向量 x_1，x_2，…，x_m 的 m 个线性组合：

$$\begin{cases} z_1 = a_{11}x_1 + a_{12}x_2 + \cdots + a_{1m}x_m \\ z_2 = a_{21}x_1 + a_{22}x_2 + \cdots + a_{2m}x_m \\ \cdots \\ z_n = a_{n1}x_1 + a_{n2}x_2 + \cdots + a_{nm}x_m \end{cases} \tag{1.16}$$

用矩阵表示为

$$\boldsymbol{Z} = \begin{pmatrix} z_1 \\ z_2 \\ \vdots \\ z_m \end{pmatrix} = \boldsymbol{A}x = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1m} \\ a_{21} & a_{22} & \cdots & a_{2m} \\ \vdots & \vdots & & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nm} \end{pmatrix}_{n\times m} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_m \end{pmatrix}_{m\times 1} \tag{1.17}$$

有

$$\begin{cases} \mu_z = \mathrm{E}(\boldsymbol{Z}) = \mathrm{E}(\boldsymbol{A}x) = \boldsymbol{A}\mu_x \\ \boldsymbol{\Sigma}_z = \mathrm{cov}(\boldsymbol{Z}) = \mathrm{cov}(\boldsymbol{A}x) = \boldsymbol{A}\boldsymbol{\Sigma}_x\boldsymbol{A}' \end{cases} \tag{1.18}$$

式中，μ_x 和 $\boldsymbol{\Sigma}_x$ 分别为 x 的均值向量和方差-协方差矩阵。

1.5 多元统计分析数值计算程序

多元统计分析涉及大量的矩阵计算，需要借助计算机来完成。本书提供的计算机程序都是采用 C 语言编写的，且已在 Windows 环境下 Embarcadero C++ Builder XE 集成开发环境下调试通过。读者可以通过作者建立的新浪博客（“学者学而时习之”博客地址：http://blog.sina.com.cn/MyMsaf）下载本书提供的执行程序及相关算例文件，或者通过作者的邮箱（liqinglaibook@126.com）进一步索取。

国外已经有很多用于多元统计分析计算的计算机软件[31]：比较专业的主要有 SAS 和 SPSS，可通过编程计算的专用软件主要有 S-Plus、R 软件和 MATLAB 等，甚至 Microsoft 的 Excel 软件也能完成多元统计分析的计算。与上述软件不同，读者可以通过本书提供的程序，不仅可以了解多元统计分析常用方法的计算过程，还可以将这些程序移植到其他操作系统或“云计算”平台上。

1.5.1 手工计算的思考

虽然多元统计分析一般需要借助计算机来完成，但根据本书算例所提供的方法，有足够的耐心和兴趣的读者，是可以通过手算来完成的，因为作者在编写程序的过程中，大多数算例都进行了手算校对。

本书所有算例虽然需要采用程序计算，但其中每一步计算并不一定依赖于某种特别的程序，是可以通过手算、本书提供的程序或其他软件完成的。下面提供的是两个“云计算”网址：

(1) 国外(“Online Matrix Calculator”)：http://www.bluebit.gr/matrix-calculator/

(2) 国内(“云算子”)：http://www.yunsuanzi.com/index.html

虽然这两个网络计算工具仅能提供矩阵的常用数值计算，不提供多元统计分析有关统计学的计算，但可以减轻手算的工作量。

1.5.2 程序计算的思考

现在的程序设计语言一般都采用了面向对象的编程技术(Object-Oriented Programming，缩写为OOP)，使软件具有可扩充性和可重用性，无论是微软的Windows(C++ Builder和Visual C++等)、苹果的MAC OS X(如Xcode的Objective-C、Unix的Objective-C)和网络(如Java语言)，还是iPad(如苹果的Cocoa)或苹果的iPhone手机上的程序，一般都采用了OOP技术。本书提供的程序大部分采用了这一方法，但读者不必为此烦恼，更不要被C++所困惑，因为采用这一技术的原因仅仅是为了子程序能够反复使用。

本章提供的已知特征值求解矩阵特征向量反幂法C语言程序，是在Mac OS X Yosemite 10.10环境下Xcode 6.2下调试通过的，本书其余程序的编译环境均是Windows环境下的Embarcadero C++ Builder XE。程序清单9.9.3-B的反幂法C语言程序是Windows下Embarcadero C++ Builder XE程序，作者已经将其改编成Mac OS X下Xcode的Objective-C程序(见程序清单1.6.3-B)。通过对比这两个程序，读者可以发现：这个反幂法程序几乎不需要修改，就可以在不同系统上编译通过。

虽然本书提供的程序均是C++，但为了若干年后C++源代码不过时，作者做了以下几个方面的努力：

1) 将部分Fortran 77程序改编成具有ANSI C标准的C语言程序

现已证实C语言具有强大的生命力，但有些语言的程序现在已经过时了。例如文献[50]所附的大量经典程序，采用的是我国20世纪60年代初自行研制的BCY程序设计语言(“编译程序语言”汉语拼音“Bianyi Chengxu Yuyan”的缩写)编写的，现在已经很难再运行在今天的计算机上了。因此，作者将其他语言的程序改编成现在的C语言程序。作者改编的主要是FORTRAN程序，程序清单9.9.3-B和程序清单1.6.3-B的InversePower反幂法子程序来自文献[17]的FORTRAN 77程序。

然而，将FORTRAN 77等语言编写的程序转化成C语言程序是件困难的事情。比如，FORTRAN语言和C语言关于数组的定义和用法有很大的不同：

$$A[1][1]_{\text{Fortran77}} = A[0][0]_{\text{C++}}$$

C 语言的变量一般是从 0 开始的，而 FORTRAN 77 是从 1 开始的(数学公式的定义一般也是这样)，当循环嵌套很多时，程序改编就十分困难。

2) 本书提供的 C 语言代码，是数学公式的 C 语言程序算术表达式

例如一个合法的 C 语言语句："c=a+b;"，这实际就是数学公式"$c=a+b$"的 C 语言程序的算术表达方式，换成其他语言也应该一样。

3) 数组的表示方法

本书程序尽量采用二维数组 A[][]来表示二维数据，增加程序的可读性，因为 C 语言中的二维数组中的元素是按行来存放的，这符合矩阵的使用习惯。一维数组 a[]也可以用来存储矩阵元素，虽然可以分配使用更大的内存，但对于数值算法来讲，每次需要计算矩阵元素存放的内存地址，可读性差，调试也容易出错，仅在部分程序中使用。

4) 减少界面菜单等影响数值计算的 C++代码的使用

由于做了上述努力，使本书提供的 C 语言程序尽可能地不依赖于系统，可以方便改编和移植。

1.5.3 常用的计算机数值计算方法

多元统计分析的计算除了涉及多元统计分析相关的统计学专用算法外，还涉及常用的计算机数值算法，主要有线性代数方程组的求解和实矩阵特征值及对应特征向量的求解。经过这些数值算法得到的结果，还需要做进一步处理后才能用于多元统计分析的下一步计算，因此了解这些算法是必要的。

多元统计分析涉及的线性代数方程组的求解和实矩阵特征值及对应特征向量的求解，都可以通过本书提出的 Schmidt 正交化 QR 分解法求解。这两部分的内容，见本书第 8 章和第 9 章。

1.6 苹果 Mac OS X 下的反幂法 C 语言源程序

本程序用于说明：本书提供的 Windows 下的 C++程序，可以方便地移植到其他系统。作者在编写和调试下面的程序前，没有在苹果系统下编写过计算机程序。

已知特征值求解特征向量的反幂法 C 语言源程序，没有封装在自编库文件中。程序清单 1.6.3-B 是反幂法子程序，程序清单 1.6.3-A 调用该子程序的主程序(注意苹果系统的 Objective-C 的主程序使用 main.m 作为文件名，而 C++则采用 main.CPP)，已在 Xcode 6.2 下调试通过。

1.6.1 程序组成

程序清单 1.6.3-A 就是完整的主程序源代码，需要调用反幂法子程序(见程序清单1.6.3-B)。

1.6.2 数据结构

程序中的变量是全局变量，见 9.9 节的说明。

InversePower 子程序来自文献[17]，原文献为 FORTRAN 77 程序，已经改编成苹果系

统 Mac OS X Yosemite 10.10 环境下的 Objective-C 程序，并在 Xcode 6.2 集成编译环境下调试通过。Windows 环境下的 Embarcadero C++ Builder XE 的程序见程序清单 9.9.3-A 和程序清单 9.9.3-B。

有关 Xcode 6.2 的 Command Line Tool 模板下的编程和调试见文献[68]，程序 main.m 运行后在输出窗口(终端：terminal)显示如下：

```
-$-$-$-$-$-$-$-$-$-$-$-$-$-$-$-$-$-$-
已知特征值计算对应特征向量的反幂法程序
1. 待求实矩阵
    1.000 000    2.000 000   0.000 000
    2.000 000   -1.000 000   1.000 000
    0.000 000    1.000 000   3.000 000
2. 通过反幂法计算的未单位化特征向量(行向量)
No.1 个特征值(精确值)=   2.000 000
特征向量(未单位化)
    1.000 000   0.500 000   -0.500 000
▼共迭代 4 次
No.2 个特征值(精确值)=  -2.372 281
特征向量(未单位化)
    -0.593 070   1.000 000   -0.186 141
▼共迭代 5 次

No.3 个特征值(精确值)=   3.372 281
特征向量(未单位化)
    0.313 859   0.372 281   1.000 000
▼共迭代 4 次
3. 单位化后特征向量(列向量)
      0.816 497   -0.503 692   0.282 184
      0.408 248    0.849 295   0.334 710
     -0.408 248   -0.158 089   0.899 078
-$-$-$-$-$-$-$-$-$-$-$-$-$-$-$-$-$-$-$-$-$-$-$-$-$-$-$-$-
Program ended with exit code: 0
```

1.6.3 已知特征值求解特征向量的反幂法 C 语言源程序（苹果 Mac OS X Yosemite 10.10）

程序清单 1.6.3-A：已知特征值主程序

```
//苹果 Mac OS X Yosemite 10.10 操作系统,Xcode 6.2 集成编译环境
//main.m
//Created by liqinglai on 15/2/17.
//Copyright (c) 2015 年 liqinglai. All rights reserved.

#import <Foundation/Foundation.h>
```