

包含大量
算法实例

Visual Basic 6.0

数理统计实用算法

李鸿吉 编著

大量精彩算例
可以举一反三

- ▶ 入门学者的自学通
- ▶ 专业人士的备忘录
- ▶ 科技人员的进阶梯

 科学出版社
www.sciencep.com

内 容 简 介

本书通过大量的算例,系统地介绍了 Visual Basic 数理统计实用算法,内容主要包括数据处理部分、图形部分和数理统计应用程序部分。本书在结构和内容的编排上注重深入浅出、循序渐进,不仅介绍应用程序,而且还交代开发过程,并避免算法的繁琐推导,同时本书配有光盘,分章节提供了源程序。

本书可以作为大专院校、培训班的教学参考书,也可以作为有 Visual Basic 开发经验的专业人士的提高读物和工具书。对于承担项目的科研人员、教学人员以及研究生,在将科技成果软件化时本书是很有实用价值的参考资料。

图书在版编目(CIP)数据

Visual Basic 6.0 数理统计实用算法/李鸿吉编著.—北京:科学出版社, 2003

ISBN 7-03-012128-7

I.V... II.李... III.BASIC 语言—程序设计 IV.TP312

中国版本图书馆 CIP 数据核字(2003)第 076630 号

策划编辑:赵卫江/责任校对:都 岚

责任印制:吕春珉/封面设计:王 浩

科 学 出 版 社 出 版

北京东黄城根北街16号

邮政编码:100717

<http://www.sciencep.com>

新 蕾 印 刷 厂 印 刷

科学出版社发行 各地新华书店经销

*

2003年9月第 一 版 开本:720×1000 1/16

2003年9月第一次印刷 印张:55 3/4

印数:1—4 000 字数:1 078 000

定价:85.00 元(含光盘)

(如有印装质量问题,我社负责调换(环伟))

前 言

在科学研究、生产实践、社会调查、军事活动和经济分析中，时时刻刻都会遇到各种类型的数据。很难设想有哪个领域与数据无关。数据无所不在，数据无时不在。

没有数据就没有发言权。数据是认识事物内在规律、研究事物之间关系和预测事物发展的基础。例如，在股市大厅里，股民朋友仰望大屏幕上不断跳动的股票数据，内心在揣摩着哪些股票该买进，哪些股票该卖出。套用一句俏皮话“分，分，学生的命根”，可以说，“数据，数据，股民的命根”。

南唐后主李煜有词曰：“剪不断，理还乱。”其实绝大多数数据就和李煜的心情一样，可以用“乱”来形容。要想从乱如麻的庞大数据堆里找出实质性的东西，做出可靠的推断，就需要对数据进行去粗取精、去伪存真的整理和检验。股市中的风云人物都是能够用跳动的数据去预测后势的强者。

数据处理是各行各业所面临的共同课题。在自然科学和技术领域里，数据处理是日常工作。今天我们能够准确地预测明天、后天的天气，那是无数个计算机处理气象数据的结果。无穷的数据，永远的数据处理，这就是我们的世界。

本书的宗旨

在数据处理中，数理统计是带有基础性的一个重要方面。

计算机与数理统计有不解之缘。早期的计算机就承担了大量的数理统计任务，成为计算机数据处理的一大支柱。为了避免应用程序重复开发，使广大用户可以应用现成的程序，有些学者很注重数理统计应用程序包的开发。例如，1985 年曾推出包括 28 个应用程序在内的“多元统计分析程序包”。

计算机硬件的发展迅猛异常，在不到半个世纪的时间里，计算机经历了从电子管、晶体管到大规模集成电路几个阶段，微机更是从 286、386、486、586 发展到奔腾系列。操作系统从 DOS、Windows 3.x 到 Windows 95、Windows 98、Windows 2000 和 Windows XP 等，也是不断地更新换代。相对于计算机硬件和操作系统的更新，数理统计应用程序的更新远远滞后，前面提到的“多元统计分析程序包”是在 DOS 下运行的，至今仍在运行。

对用户而言，如果经常要从 Windows 系统切换到 DOS 系统去运行某些应用程序，是一件很令人烦恼的事。为了使用户摆脱繁琐的操作过程，充分享受最新操作系统和先进的可视化应用程序开发工具带来的种种便利，需要对以前的应用程序进行升级换代和界面更新，这是一件应该必须做、及早做的事。这正是本书的宗旨。

本书的主要内容

本书共 19 章，前 18 章分成以下三大部分：

- 第 1、2、3、6、7 章是数据部分。数理统计是要靠数据说话的，离开数据，数理统计就无从谈起。在这部分中，建立了数据录入和编辑过程，提供一个包括数据库与数据文件相互切换的通用数据库软件，这些应用程序为用户进行数据整理并建立用于运行的数据文件带来极大的方便。第 6 章和第 7 章介绍用于作图的一般性数据处理，例如光顺和插值。
- 第 4 章和第 5 章是图形部分。如果数据处理的结果还需要用手工描绘，那是一件令人不开心的事。这部分包括绘制曲线、等值线图和立体图，也可以制作图表。
- 从第 8 章一直到第 18 章，都是数理统计应用程序部分，其中包括随机变量的分布函数、假设检验、回归分析、相关分析、趋势分析、方差分析、点群分析、判别分析等，是本书的主干。

为了将上述三部分有机地联系在一起，在第 19 章建立了一个“总界面”，通过“总界面”用户可以很方便地实现不同应用程序之间的切换。

本书的特点

- 应用程序开发时，使用的操作系统是 Windows 2000 和 Windows XP，使用的语言是 Visual Basic 6.0。
- 本书不仅把应用程序介绍给读者，而且还向读者交代开发过程。对开发过程中所涉及的 Visual Basic 6.0 的窗体、控件对象、函数等，都根据程序开发进程逐步进行介绍。这些内容对读者加深理解，并进行自主开发都是必要的。
- 通俗易懂、循序渐进是本书坚持的原则之一。前 3 章所涉及的 Visual Basic 6.0 都是最基本的，后面章节随着开发进程逐渐深入。在介绍程序源代码时添加大量注释，这样可以提高代码的可读性。
- 本书在介绍算法时避免繁琐的推导。如果读者对推导感兴趣，可以参阅本书所列的参考文献。
- 算例贯穿本书的始终。算例可以带给读者感性认识，同时算例也是对应用程序的使用方法的直观说明。

本书的读者对象

本书对读者的 Visual Basic 水平没有确切要求。Visual Basic 水平较高的读者，可以跨过有关 Visual Basic 用法的介绍，跳跃式阅读。

本书面向使用数理统计的广大读者：

- 对那些只想尽快得到数理统计结果的读者，本书是“数理统计应用程序包”。只要建立了相关的数据文件，点点鼠标就可以取得立杆见影的计算结果，如果需要的话，作图也是一件轻而易举的事。
- 对那些想进一步使用 Visual Basic 进行数理统计算法及其他数据处理算法开发的读者，本书是“自学通”。从建立数据文件、形成数据库应用程序、开发数理统计算法到编制绘图程序，读者都可以从本书找到对应的有参考价值的算法。

如何使用本书

- 如果读者只想尽快得到数理统计的结果，建议重视第 2 章，了解如何建立数据文件，再找到需要的算法章节，特别是通过算例看一看算法对数据文件格式有没有特殊要求，然后再整理数据并建立数据文件。以后的事情是按照算例的步骤点点鼠标就可以了。在对应用程序用法有了一些了解之后，使用第 19 章的“总界面”会更方便。
- 如果读者已经有了自己的输入、输出过程，只是想了解方法，建议这些读者特别重视大多数程序都有的 BAS 模块 modMethod，这是用过程构筑的方法模块，本书对每个过程的调用格式都有详细说明。但这也不绝对，有少量程序将计算语句放到窗体 frmCalculate。
- 如果读者想进行自主开发，又对 Visual Basic 不太熟悉，那么从头开始仔细阅读还是有必要的，同时可以一步一步考虑所读的内容如何纳入自己的体系。如果感觉本书对 Visual Basic 的介绍不够详细，可以参阅笔者近年编写的《Visual Basic 6.0 编程方法详解》（以下简称《详解》）和《Visual Basic 高级编程技术》（以下简称《技术》），两本书都是由科学出版社出版的。

如何使用光盘

建议读者特别重视光盘的作用。

本书的光盘分章节提供源程序，都与书的文字部分相对应。

可以将光盘程序复制到硬盘，也可以直接运行光盘。光盘中也包括本书算例所使用的数据，如果能够按书中算例的步骤重复算法，可以取得事半功倍的效果（直接运行光盘时，如果需要保存数据，要注意路径设置）。

不管是硬盘，还是光盘，若要运行源代码，需要进入 Visual Basic 6.0 的开发环境。此外，本书的程序没有打包，若要运行 EXE 文件，也需要先在计算机内安装 Visual Basic 6.0（这时无需进入开发环境）。

感谢

李玉书研究员和万永革博士参加部分工作。

如果没有全家人的热情鼓励和鼎立支持，本书是不可能完成的。

一年以来，笔者不敢懈怠，全身心地投入到本书的编著中，力求完美。但由于水平有限，书中错漏在所难免，恳请专家和广大读者指正。

作 者

目 录

第1章 数据录入	1
1.1 数组	1
1.1.1 数组的声明和应用	1
1.1.2 动态数组	2
1.2 过程	4
1.2.1 子程序	4
1.2.2 函数	5
1.2.3 引用和传值	5
1.2.4 Sub Main()	6
1.3 BAS 模块	7
1.4 窗体	8
1.4.1 窗体的定位	8
1.4.2 窗体事件	10
1.4.3 多窗体和启动窗体	11
1.4.4 窗体语句和方法	12
1.4.5 多文档界面	13
1.5 控件	14
1.5.1 共同属性	15
1.5.2 通用语句和方法	16
1.5.3 录入程序所使用的控件	16
1.5.4 控件数组	22
1.5.5 控件集合	23
1.6 图形容器	24
1.6.1 图形容器属性	24
1.6.2 图形容器方法	25
1.7 MsgBox 函数和语句	27
1.8 InputBox 函数	30
1.9 用文本框数组实现数据录入	30
1.10 用网格实现数据录入	34
1.11 小结	38
第2章 数据文件	39
2.1 单选钮和复选框	39

2.1.1	单选钮.....	40
2.1.2	复选框.....	40
2.2	文件的类型.....	40
2.3	顺序文件	41
2.4	随机文件	43
2.5	二进制文件.....	45
2.6	文件函数和语句	45
2.6.1	文件函数和语句表.....	45
2.6.2	文件函数和语句的应用.....	46
2.7	文件系统对象.....	48
2.7.1	文件系统对象加载.....	49
2.7.2	文件系统对象的建立和操作.....	49
2.7.3	文件系统对象的层状结构	49
2.8	文件控件	54
2.8.1	驱动器列表框.....	55
2.8.2	目录列表框.....	55
2.8.3	文件列表框.....	55
2.8.4	文件控件的组合使用	56
2.9	公共对话框.....	57
2.9.1	公共对话框的类型设置.....	57
2.9.2	公共对话框的“打开/另存为”选项卡.....	58
2.9.3	公共对话框的“颜色”选项卡.....	61
2.9.4	公共对话框的“字体”选项卡.....	61
2.9.5	公共对话框的“打印”选项卡.....	62
2.9.6	公共对话框的“帮助”选项卡	63
2.10	数据文件_录入	63
2.11	数据文件_编辑.....	78
2.12	小结	87
第3章	数据库.....	89
3.1	列表框.....	89
3.1.1	列表框属性.....	89
3.1.2	列表框方法.....	91
3.1.3	列表框事件.....	91
3.2	菜单.....	91
3.2.1	菜单编辑器.....	91

3.2.2	下拉式菜单.....	94
3.2.3	弹出式菜单.....	99
3.2.4	菜单项规划表.....	102
3.2.5	菜单数组.....	103
3.2.6	近期文件菜单.....	107
3.3	数据库编程方法.....	114
3.3.1	数据库结构.....	114
3.3.2	数据控件.....	115
3.3.3	绑定控件.....	119
3.3.4	多表关联.....	120
3.3.5	结构化查询语言.....	123
3.3.6	DAO 对象编程.....	127
3.4	通用数据库管理.....	142
3.4.1	通用数据库管理程序的功能.....	143
3.4.2	通用数据库管理程序的组成.....	143
3.4.3	通用数据库管理程序的代码.....	145
3.4.4	通用数据库管理程序的运行.....	187
3.5	小结.....	196
第 4 章	图表.....	197
4.1	图表控件.....	197
4.1.1	图表控件的加载.....	197
4.1.2	图表控件的属性.....	197
4.1.3	图表控件的类型.....	198
4.1.4	图表控件的数据.....	198
4.2	图表绘制.....	200
4.2.1	图表绘制程序的功能.....	200
4.2.2	图表绘制程序的组成.....	200
4.2.3	图表绘制程序的代码.....	201
4.2.4	图表绘制程序的运行.....	205
4.3	小结.....	207
第 5 章	图形.....	208
5.1	控件对象在窗体上的移动.....	208
5.2	组合框.....	210
5.3	曲线图.....	211
5.4	等值线图.....	222

5.5	彩色等值线图	240
5.6	立体图程序	251
5.7	网状立体图程序	267
5.8	小结	280
第 6 章	曲线数据处理	282
6.1	滑动平均	283
6.1.1	滑动平均计算公式	283
6.1.2	滑动平均算法	283
6.2	曲线插值	293
6.2.1	曲线插值计算公式	293
6.2.2	曲线插值算法	295
6.3	小结	313
第 7 章	曲面数据处理	314
7.1	曲面数据处理原理	314
7.1.1	近点按距离加权平均原理	314
7.1.2	加权最小二乘拟合的基本原理	315
7.2	曲面数据处理算法	316
7.3	小结	331
第 8 章	描述统计	332
8.1	描述统计内容	332
8.1.1	集中趋势	332
8.1.2	离散程度	333
8.1.3	偏度和峰度	334
8.2	描述统计算法	335
8.3	小结	344
第 9 章	指数	345
9.1	物价指数	345
9.1.1	几何平均指数	345
9.1.2	综合指数	345
9.2	指数算法	347
9.3	小结	354
第 10 章	频数分布	355
10.1	频数和累积频数	355
10.2	频数分布算法	356
10.3	小结	371

第 11 章 随机变量的分布函数	372
11.1 正态分布	374
11.1.1 正态分布函数	375
11.1.2 正态分布分位数	379
11.2 二项分布	381
11.3 泊松分布	389
11.4 Γ 分布	396
11.5 χ^2 分布	397
11.5.1 χ^2 分布函数	397
11.5.2 χ^2 分布分位数	400
11.6 B 分布	404
11.7 t 分布	405
11.7.1 t 分布函数	405
11.7.2 t 分布分位数	409
11.8 F 分布	415
11.8.1 F 分布函数	416
11.8.2 F 分布分位数	419
11.9 小结	424
第 12 章 假设检验	425
12.1 参数假设检验	427
12.1.1 总体平均数假设检验	427
12.1.2 总体方差假设检验	443
12.1.3 总体比例假设检验	459
12.2 分布假设检验	471
12.2.1 正态分布假设检验	472
12.2.2 二项式分布假设检验	473
12.2.3 泊松分布假设检验	473
12.2.4 分布假设检验算法	474
12.3 小结	498
第 13 章 方差分析	499
13.1 单因素等次方差分析	499
13.1.1 单因素等次方差分析的统计假设	499
13.1.2 单因素等次方差分析的离差和自由度	500
13.1.3 单因素等次方差分析的 F 检验	501
13.1.4 单因素等次方差分析算法	501

13.2	单因素不等次方差分析	513
13.2.1	单因素不等次方差分析的离差和自由度	513
13.2.2	单因素不等次方差分析算法	514
13.3	双因素方差分析	525
13.3.1	双因素方差分析的离差和自由度	525
13.3.2	双因素方差分析的统计推断	526
13.3.3	双因素方差分析算法	527
13.4	双因素交错方差分析	543
13.4.1	双因素交错方差分析的离差和自由度	543
13.4.2	双因素交错方差分析的统计推断	545
13.4.3	双因素交错方差分析算法	545
13.5	小结	559
第 14 章	回归分析	560
14.1	一元线性回归	560
14.1.1	一元线性回归的数学模型	560
14.1.2	一元线性回归方程的系数估计	561
14.1.3	一元线性回归的显著性检验	562
14.1.4	利用一元线性回归方程预测、平滑	563
14.1.5	一元线性回归算法	563
14.2	多元线性回归	582
14.2.1	多元线性回归的数学模型	582
14.2.2	多元线性回归方程的系数估计	583
14.2.3	多元线性回归的显著性检验	584
14.2.4	利用多元线性回归方程预测、平滑	585
14.2.5	多元线性回归算法	585
14.3	逐步回归	606
14.3.1	逐步回归的基本思想	607
14.3.2	逐步回归实现步骤	608
14.3.3	逐步回归实现说明	610
14.3.4	逐步回归算法	610
14.4	多项式回归与多项式逐步回归	633
14.4.1	多项式回归算法	633
14.4.2	多项式逐步回归算法	643
14.5	一元非线性回归与非线性逐步回归	662
14.5.1	一元非线性回归	662

14.5.2	非线性逐步回归	683
14.6	小结	702
第 15 章	相关分析	703
15.1	简单相关分析	703
15.1.1	简单相关系数的意义	703
15.1.2	简单相关系数的特点	704
15.1.3	简单相关系数的检验	705
15.1.4	简单相关系数算法	705
15.2	复相关分析	715
15.2.1	复相关系数	715
15.2.2	偏相关系数	716
15.2.3	复相关系数与偏相关系数算法	717
15.3	小结	735
第 16 章	趋势分析	737
16.1	滚动条	737
16.2	二元多项式回归分析	739
16.2.1	二元多项式回归方程的建立	739
16.2.2	二元多项式回归方程的检验	740
16.2.3	二元多项式回归方程算法	741
16.3	二元多项式逐步回归算法	774
16.4	小结	802
第 17 章	点群分析	803
17.1	相似统计量	803
17.1.1	距离系数	803
17.1.2	夹角余弦	804
17.1.3	相关系数	804
17.2	点群分析的数据预处理	805
17.2.1	标准差标准化	805
17.2.2	极差标准化	805
17.2.3	极差正规化	805
17.3	点群分析的聚类方法	806
17.4	点群分析算法	806
17.5	小结	827
第 18 章	判别分析	828
18.1	两组判别分析	828

18.1.1	两组判别分析的判别方程.....	828
18.1.2	两组判别分析的判别指标及显著性检验.....	829
18.1.3	两组判别分析算法.....	830
18.2	逐步判别.....	847
18.2.1	逐步判别基本原理.....	847
18.2.2	逐步判别计算步骤.....	848
18.2.3	逐步判别算法.....	849
18.3	小结.....	865
第 19 章	总界面.....	866
19.1	Shell 和 App.....	866
19.1.1	Shell 函数.....	866
19.1.2	App 对象.....	867
19.2	总界面的开发步骤.....	867
19.3	总界面算法.....	869
19.4	小结.....	875
主要参考文献		876

第 1 章 数据录入

利用 Windows 的“记事本”(Notepad)可以录入纯文本数据文件。这种方法的优点是不必编程,省事。但有两个缺点:

- 不具有引导性,用户必须熟悉并记住数据文件的格式。
- 用户界面不够友好,数与数之间没有特别明显的分界。假如能够将录入的每个数据都放在屏幕上的网格中,看起来会更清楚,改起来也容易。

充分利用 Visual Basic (以下简称 VB) 的优势,编制用于录入数据的专用程序是最佳选择。

本章的主要内容是介绍用文本框数组和网格实现数据录入的方法。在进入正题之前,先结合 VB 的特点,对两个程序所涉及到的程序设计语法、控件等方面的知识做简要说明。

1.1 数 组

假如计算 100 个整型变量的和,这时就需要定义 100 个整型变量,很繁琐,也许能勉强做到。那么 1000 个怎么办? 10000 个又怎么办? 这时必须用数组。

数组是保存相同数据类型的多个变量的存储空间。数组具有整体性、有序性和易操作性,在处理大量的数据时,变量无能为力,而用数组则可以迎刃而解。

今后,我们在进行数据处理时,几乎时时刻刻都和数组打交道。

1.1.1 数组的声明和应用

1. 数组的命名

在 VB 中,数组的名称不能超过 255 个字符,由字母、数字、下划线组成,必须以字母打头,数组名称不能是 VB 的关键字(例如,For、Next、Me 等)。例如, intX、sng123、dblArray 等都是合法的数组名称,如果以 123sng、_ABC、Then (VB 的关键字)等作为数组名称则是非法的。

2. 数组的下界和上界

数组元素的个数由下界和上界确定。

VB 的缺省下界是 0,例如, Dim intX(99)As Integer 定义了下界为 0、上界为

99 的 100 个整型变量的数组。如果不习惯于数组从 0 开始,这时可以用 `Dim intX(1 To 100) As Integer` 设定数组的下界为 1。

如果在模块或窗体的通用声明部分加上 `Option Base 1`, 这时数组的缺省下界不再是 0, 而是 1, 在这种情况下, `Dim intX(1 To 100)As Integer` 可以简单地写成 `Dim intX(100)As Integer`。

3. 多维数组

`Dim intXY(2,99) As Integer` 定义了一个二维整型数组, 分别从 0 到 2, 从 0 到 99。自然, 这类数组也可以用 `Dim intXY(1 to 3,1 To 100)`来定义。

VB 规定, 数组的最大维数是 60。

4. 数组定义语句类型

除了 `Dim` 之外, VB 还可以用 `Public`、`Private` 等定义数组。不同的定义语句有不同的应用范围和不同的作用域, 分述如下。

- **Public:** 用于 BAS 模块。用 `Public` 所声明的数组的作用域为整个程序。
- **Private:** 用于模块或窗体。其作用域分模块级或窗体级。
- **Dim:** 用于模块、窗体或过程。其作用域分模块级、窗体级或过程级。
- **Static:** 用于过程体内。用 `Static` 定义的数组在出入过程时, 数组的值不改变, 即不被初始化为 0。
- **ReDim:** 用于过程体内。为动态数组分配内存空间。

5. 数组的应用

假定有关的变量和数组已经声明, 并已经为数组 `intX` 赋值, 使用循环语句可以很容易地实现求 1000 个数据的和:

```
IntS = 0
For intI = 1 To 1000
    IntS = intS + intX(intI)
Next intI
```

1.1.2 动态数组

1. 数组的类型

在 VB 中有两种类型的数组。

(1) 固定数组

下界和上界都不变。

(2) 动态数组

下界或上界可以改变。

2. 使用动态数组的必要性

如果事先能肯定有多少数据需要处理,而且以后再次处理时数据量也不变,使用固定数组也无妨。事实上,对绝大多数课题而言,事先并不能确切知道数据量大小,这次和下次需要处理的数据量也可能不同,第一次求 100 个数据的和,第二次可能求 1000 个数据的和,甚至 10000 个。若将数组卡死了,则每次都需要修改源程序,重新编译,这实在是下策。

为了避免修改源程序,一个较为常用的做法是,给数组设定充分多的维数,足够大的元素数,但这样做既浪费了空间资源,也降低了运行效率。

要想编制灵活的、通用性强的程序,使用动态数组是惟一的选择。

3. 动态数组的使用

本书在事先不能或不便规定数组大小的地方,一律使用动态数组。

本书在使用动态数组时:

- 数组一律规定为二维;
- 下界一律规定为 1;
- 从数据文件或其他途径获得上界。

4. 动态数组的声明

- 先在 BAS 模块的说明部分作没有维数和元素数的初声明,例如,

```
Public dblArray() As Double
```

- 由所获得的上界,在某一个过程体内重新定义数组,例如,

```
Private Sub cmdOK_Click()  
    intRow = Val(txtRow.Text)  
    intCol = Val(txtCol.Text)  
    ReDim dblArray(1 To intRow, 1 To intCol)  
End Sub
```

在上面的程序段中,txtRow 和 txtCol 是窗体上的文本框对象。intRow 和 intCol 是整型变量。Val 是成员函数,可以将字符串转换成数字。

经过重新定义的数组在整个应用程序内有效。

5. Preserve 关键字

VB 允许反复使用 ReDim 语句,可以多次改变数组的维数和每一维的元素个数,但不能改变数组的类型。如果动态数组中以前的数据需要保留,可以使用 Preserve 关键字,例如,

```
Redim Preserve dblArray(1 To intNewRow, 1 To intNewCol)
```