数字计算机上用的数学方法

第三卷

——统计方法——

上海科学技术出版社

数字计算机上用的数学方法

第 三 卷

—统 计 方 法—

〔美〕K. 安斯伦　A. 拉尔斯登　H. S. 维尔夫

中国科学院计算中心概率统计组　译

上 海 科 学 技 术 出 版 社

内 容 提 要

本书是继第一、二两卷(中译本分别于1963年、1976年出版)之后专门介绍在数字计算机上进行统计计算的一些数学方法。全书分五篇：引论，回归分析与判别分析，主分量分析和因子分析，聚类分析和模式识别，时间序列。共十五章。其中第一章是概论性的，介绍统计计算的现状，特别是和本书各章有关的多元统计计算的现状。

各章叙述的格式，除第一、二两章外，仍和第一、二两卷一样，采用大体上一致的处理格式，分成职能、数学讨论、计算过程摘要、框图、框图说明、程序和子程序、例题、机器工作时间估计、参考文献等九个项目。

本书可作为在数字计算机上从事实际计算工作者的一本参考书，也可供从事应用统计、应用数学、数值分析的读者参考。

STATISTICAL METHODS FOR DIGITAL COMPUTERS

VOLUME III of

Mathematical Methods for Digital Computers

K. Enslein A. Ralston H. S. Wilf

John Wiley and Sons, Inc., 1977

数字计算机上用的数学方法

第 三 卷

——统 计 方 法——

〔美〕K. 安斯伦 A. 拉尔斯登 H.S. 维尔夫

中国科学院计算中心概率统计组 译

上海科学技术出版社出版

(上海瑞金二路 450 号)

新华书店上海发行所发行 上海商务印刷厂印刷

开本 787×1092 1/16 印张 23.25 字数 560,000

1981 年 7 月第 1 版 1981 年 7 月第 1 次印刷

印数 1—10,000

书号：13119·848 定价：(科四) 2.15 元

序

统计计算是从事科学计算部门的重要计算任务之一. 在这套丛书的第一卷里，有四章关于统计计算的文章，但用现在的标准来看，它们都是一些过时的或极为粗糙的统计计算方法. 在第二卷里，没有再包括统计计算方面的文章. 所以，原编者(A. 拉尔斯登，H. S. 维尔夫)认为，对这样一个重要领域，现在应该专门编辑一本书. 我们邀请 K. 安斯伦作为本书的第三个编者. 在统计计算领域里，他是专家，有着丰富的实际经验.

我们的宗旨是用和本丛书前两卷大致相同的篇幅，尽可能多的包括统计计算领域中最重要的和最常用的一些方法. 我们尽力保持前两卷中的一些主要特点，特别，在每篇文章中，包含大致相同的论题，并尽可能地保持相同的格式.

象第二卷那样，如果一个算法的程序不太长，就给出实际的程序. 所有这些程序都是用 Fortran 编制的. 在多数情况下，Fortran 是统计计算常用的一种算法语言. 在第一章中，给出本书所讨论的全部问题的指南.

为了很好理解本书各章的内容，读者至少应具备使用初等统计的一些经验和一定的数学水平. 但是，把本书作为计算方法，框图，程序和参考文献的概要来读，也是有益的.

本书各章的作者都是一些著名的专家，在他们各自从事的领域内有过重要的贡献. 他们利用这次著书的机会，加深了对这些问题的讨论. 在这里，我们对本书各章的作者表示感谢. 我们希望读者从他们的经验中，象我们一样，从中得到教益.

K. 安斯伦
A. 拉尔斯登
H. S. 维尔夫
1976. 4.

引　言

一、目　的

在《数字计算机上用的数学方法》的第一卷中，和统计方法有关的共四章：多重回归分析，因子分析，自相关及谱分析和方差分析．在1960年，当第一卷出版时，这四章内容在一定程度上反映了当时统计计算的范围．但近15年来，统计计算有了许多重大发展，许多新的有效的统计计算算法得到了广泛应用．本卷中给出的十五章统计算法，只是这些最重要、最有用的统计算法的一部分．由于今日统计计算范围的广泛，不能指望用一本书包揽全局，所以这十五章只能是统计计算内容的一个代表．本卷书中未能包括的一些重要的统计计算领域，将在第一章里粗略提到．

二、各章的格式

象本丛书前两卷一样，编者想尽可能使更多的章节具有一致的格式．在这一卷中，第一章是概论性的，不取这种标准格式．第二章，由于它的内容，也无法具有这种标准格式．从第3章到第15章，除个别例外，都包括下述一些内容：

职能

所要描述的方法的职能是什么？各章的作者力图在他所写的一章中对所考虑的特殊问题，给出一个简洁、正确的提法．

数学讨论

就本章包括的范围，作者力图给出需要求解问题的完整的数学描述，提出求解问题的一种或数种方法．在这里，读者可以找到有关的数学定理及其证明或者和证明有关的参考文献．如果有适用的误差分析方法，读者可找到误差分析应用好坏情况的讨论．读者在这里还会找到和其它有关方法的比较及引用的有关文献．

计算过程摘要

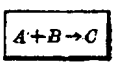
当统计讨论和方法推导交叉进行时，读者常常会弄不清楚求解这一问题的精确步骤．所以，在这一节中，读者可以找到前面导出方法的一种开"药方"式叙述的解题步骤，即首先做这个，其次做那个，等等．

框图

这里给出的框图，按所含资料的一般性和允许的篇幅，尽可能详细．

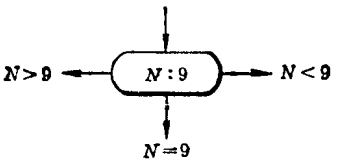
在框图中，框的类型用得尽可能少．所有框图以标有"启动"的一框开始，而以标有"停机"的一框结束．用到的其他类型的框有：

1．叙述框



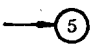
此框叙述框内的操作在通过这一框时执行．

2．检查框



这种框有一个入口和三个可能出口．三个出口，按照大于、小于或等于选择一个．当仅有两个可能出口时，按照框内问题的成立与否选择一个是或否的出口．

3．远接框



这圆圈指明逻辑控制转移到另一个标有同一数字的圆圈．

框图说明

给出框图中逐框的说明，以便读者有所依从．

程序和子程序

由于篇幅的限制，这里只给出一些重要的 Fortran IV 子程序．在任何情况下，都列出所需要的标准子程序．在第一章中，列出了一些单位的名称，从它们那里可以得到这些程序．

例题

做出一个有代表性的例题，使读者对解题的实际过程有所了解．

参考文献

列出文中引用的一些参考文献，用方括弧中的数字表示引用参考文献的序数．

目　录

第一篇

引　论

1 统计计算园地导游

K. Enslein　　张建中 译

1. 引　言

这套丛书中的前两卷，主要讨论数字计算机上用的数学方法，只是偶尔涉及到数字计算机上用的统计方法．在这一卷中，将全部讨论后者．统计计算的范围极其广泛，不可能在一本书中进行非常全面的论述．因此，编者面临一种非常困难的选择：在这样一本书中，应该选入哪些统计方法，又应排除哪些．本书选入的一些方法，从两方面来讲，即从它们在统计领域中的重要性以及它们在计算机上应用的广泛性来讲，我们认为它们都是重要的．

在这一章中，将回顾一下统计计算的现状，特别是和本书各章有关的多元统计计算的现状．关于多元统计分析发展的近期趋势，读者可参考 C. R. Rao 的文章[1]．

作者原想对本书各章给出的以及尚未编入的一些统计方法，给出一种描述性的介绍．但经认真考虑后，作者决定从问题求解的观点出发组织这一章，并尽可能多用一些图表进行说明．这样做对读者可能更为有益．因此，在这一章中，文字较少，相对说来图表却较多．

2. 方　法

在确定应用哪些统计分析方法求解一个给定问题时，从问题求解者的观点来看，并没有太多的统计问题需要确定解法，更多的是如何使用一个给定的统计方法．换言之，整个取决于问题的求解者想要做些什么．在表 1 中，给出了使用多元统计分析方法的简单说明．但表 1 过于简单，不足以说明方法和目的之间的关系．所以，从不同的观点出发又给出了图 1.1．在图 1.1 中，不是一个目的对应一种方法，而是给出达到一个目的可用的一系列方法．

下面给出图 1.1 中使用的各个术语的定义，可能对读者理解该图会有一定帮助．

1．*模型构造和模型外推*：导出解释应变量方差(变异)的方程或方程组．例如，给出总体寿命，希望用一些观测参数，如生活习性、状态和遗传等[2]，解释这一变异．

2．*协变量调整*：这里比通常的理解具有更一般的意义，目的是使变量、特别是应变量标准化．经过这种方法处理，可以消去一些名为可控、实不可控的干扰因素．例如，根据从外科到精神病科各个不同科室的医生对精神病患者观测的结果，研究一种精神病药物的疗效．如果问每一个医生：病人对这种药物需要的迫切程度如何？可以发现，拿外科医生和精神病科的医生相比，外科医生会认为病人对这种药物需要的迫切程度要低一些，而其他医生和实习医生的回答，往往会介于二者之间．这时，协变量调整的目的，就是调整每一个观测者对病人的迫切愿望的反映，使得从不同观测者得到的观测结果之间可以互相比较．为此，医生的一些特点，如他们的专长、实际经验的多寡等，就必须加以考虑．

表1 在数字计算机上应用统计方法的限制和特征

特征 \ 章别	3 子集选取	4 逐步回归	5 逐步判别分析	6 组合方法	7 因子分析(Jöreskog方法)	8 因子分析(Minres方法)	9 唯一旋转	10 多元方差和协方差分析	11 谱系分类	12 多维尺度转换	13 ISODATA聚类	14 FFT	15 预报
中央存储器容量													
小(20K字)	V	20V	V×G=80	V×G=50	V×F=80	V×F=100	V	N×V×F=200	40N	N×F=40	N×V=2000	100I	100I
中(60K字)		100V	V×G=200	V×G=100	V×F=200	V×F=300		N×V×F=400	100N	N×F=100	N×V=50000	500I	500I
大(≥100K字)		400V	V×G=600	V×G=400	V×F=600	V×F=800		N×V×F=1000	400N	N×F=400	N×V=200000	2000I	2000I
二进制字长													
32	P	P	P	P?	P	P	P	P		P	P	P	P
36	P?	P	P	P?	P	P	P	P?		P?		P?	P?
48～60				P?	P-		P?						
计算速度													
算术浮点加 <1微秒	70V					V×F=400	50V	N×V×F=200		N×F=100			
算术浮点加 ≥1微秒	30V	50V	V×G=200	V×G=150	V×F=300	V×F=100							
逻辑运算									N				
图形输出	无	有	有	有	无	无	有	无	有?	有?	无	有	有
"代价"	V		V×G×N	V×G×N	V×F	V×F	V×F	N×V×F	N	N×F	N	I	I
对应解的唯一性	不唯一	不唯一	不唯一	不唯一	不唯一	唯一	不唯一	?	不唯一	不唯一	不唯一	不唯一	不唯一
优化	是	否	否	是	是	是	是?	是	否	是?	是?	是?	是?
辅助存储		需要	需要	需要				需要	有帮助		有帮助		

F=因子数 I=区间数 P=准确 G=组数 N=观测次数 V=变量个数 ?=不能确定

3．分类：即对象分组．在聚类分析和模式识别部分的编者前言中，对这一专题将进行更深入的讨论．这里，我们把有参考可循的分类称为有师可学的分类，把无参考可循的分类称为无师可学的分类．

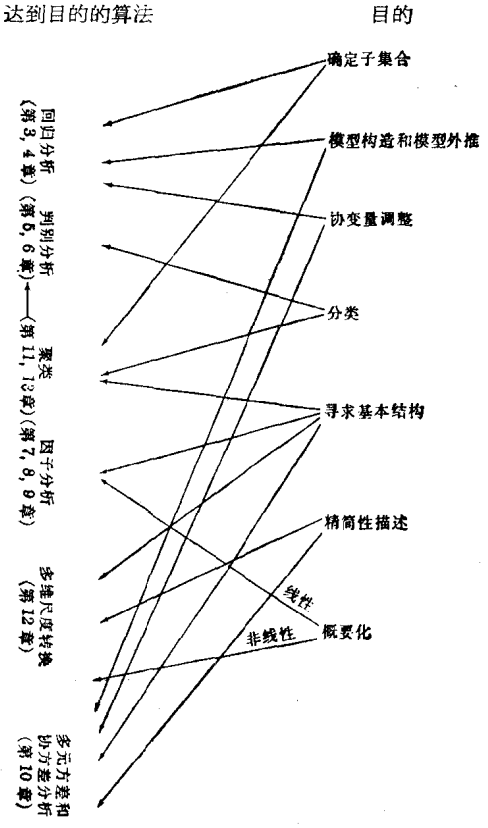


图1.1 数据分析方法说明
（对时间序列分析另作说明）

聚类：可以看作是无师可学的一种分类方法．在判别分析中，预先给出了分类结果，希望找到能把各类分开的超曲面，是一种有师可学的分类方法．

4．寻求基本结构：给定一组变数或观测数据，我们要问，数据的维数是否一定要大到象已有的变数或已有的事物那么多．在已有的变数或已有的事物中，是否存在一个子集，特别是一个加权子集，能够以预先给定的误差阶数说明整个问题的复杂性？因子分析是解决这类问题的一种古典方法．近些年来，已把多维尺度转换法作为一种非线性因子分析方法使用．也可以把结构看作方差的分量，因此经常用到多元方差和协方差分析．

古典基本结构分析问题产生于个性因子分析中，读者可参考这一领域中的许多论著，特别是Harman[4]和本书的第8章及Cattell的第9章．

5．精简性描述：除要求用最小个数的变量说明整个数据的复杂性外，它和寻求基本结构有着一定的关系．例如，在生物过程中，可以观测到各类不同的症状，但大多数症状都可用很少几个与其有关的基本特征进行说明．这里再次说明它和简单结构有着清晰的关系．

6．概要化：就概要化而论，它和精简性描述有许多类似之处，只是推广到了多元总体．例如，用因子结构或其他线性、非线性的多维表示，描述一些个性特征．

后面三个术语的范围是不太容易分开的，但在求解问题时，它们有着明显的不同．因此，在这一章中，我们决定保持它们各自的本性．

3．回归分析和判别分析

a．理论和算法

自从Fisher的著作[5]出版后，这方面的理论研究有了不少进展．就回归分析来讲，大多数研究工作者都致力于“最优”变数子集的选择问题上．这一动向是由于在大量的研究中，遇到了很多变量，且多为非正交变量的实际问题而产生的．所以，在多元回归分析中，要从变量组中选取一个最优子集，能够很好地“解释”应变量的方差；在多元判别分析中，要从变量组中选取一个最优子集，能够“很好”地将已经给定的各组分开．这些方法通常分为渐增法（聚集法，向前消去法）和渐降法（向后消去法）．就渐增法而论，是指在一次计算中，根

据某种标准，如加入回归方程中的变量使应变量的剩余方差减少最大或使复相关系数增加最多等，把一个变量引入回归方程．渐降法是从所有变量都已引进回归方程开始的，根据某种标准，如减少复相关系数最小，一次消去一个或几个变量．

现在，对逐步方法能否给出最稳健(Robust)的结果有着很大的争论．聚集变数是最常用的一种算法[6](参见第4章、第5章)，有些人，如Mantel[7]，用大量令人信服的论据，论述渐降法能够给出更稳健的方程．可以进行论证的方法有岭回归或阻尼回归[8]．多年来，工程技术人员已把这种方法用于非线性最优化问题．把一个很小的阻尼因子加到相关矩阵或协方差矩阵的对角元素上去，就可以对变量进行正交化处理，也可以把这种方法推广去处理不定问题，即变量个数多于观测次数的问题．岭回归可当渐降法使用，因而进一步支持了Mantel关于渐降法较优的论述．

在判别分析中，特别在逐步判别分析问题中，原有理论都假定已分各组具有相同的协方差矩阵．最近，Wilf发展了必要的理论和算法，允许处理具有不同协方差矩阵的分组．这一方法将在本书第6章里介绍．从技巧上讲，这是一个重大发展，在某些方面，等价于如ISODATA(第13章)引入的逐段线性判别分析的方法，也类似于判别分析问题中把线性和逐段线性组合在一起的一些方法．

b．计算

稳健性

上面讨论中提到的稳健性问题，已由Tukey用称之为"Jackknifing"的方法开始着手解决．在多元回归分析或多元判别分析问题中，用消去事物或变量子集合的方法计算一些系数的估计值，再计算平均系数的整个估计．在这个领域中的主要参考资料里，[11]、[12]、[13]是比较典型的．

多重回归

在加利福尼亚大学出版的BMD[14]一书中，有最常用的多元回归分析程序．多年来使用的计算方案称为BMD02R．最近，引进了方案BMDP02R，包括选取附加变量的一些算法和比较完整的残差绘图程序．BMDP02R更强调了渐增法，在许多方面进一步补充了Eforymson的算法[6]．在名为STEPREG的过程中，大大推广、补充了渐增法．在本书的第4章中，Jennrich给出了这一算法．

渐降法并不常用，多以专用程序实现，未见提供出来．

岭回归

有不少私人的岭回归文本．公开发表的，据知只有RIDGE[15]．这类算法能自动去掉一些"弱"变量，是一种渐降法．

逐步判别分析

逐步判别分析发展的历史和逐步回归分析发展的历史十分类似．最常用的逐步判别分析程序在加利福尼亚大学出版的BMD一书中，叫做BMD07M．这是一种渐增法，在本书的第5章中介绍．

选择子集的其他一些方法

LaMotte和Hocking发展了一种名为SELECT的算法，经过比较简单的分叉与剪枝算法，计算2个，3个，…，n个变量最优子集合的复相关系数，这是对考察所有变量的可能组合并从中选出具有最大复相关组合的那类算法的"最有意义"的逼近．第3章将讨论这一算法．

4. 主成分分析和因子分析

a. 理论和算法

从历史上来看，远在快速数字计算机出现之前，为了满足心理学家的需要，主成分分析和因子分析就已经有了很大的发展. 虽然寻求主成分的一些方法早由 Hotelling[16]给出，但和旋转主成分向量的一些方法相比，仍有显著差异. 问题是要给出最优旋转的定义. 从最早把主成分向量旋转到方差最大变位上去的方法，即把主成分向量旋转到协方差矩阵残差平方和降到最小并保持因子正交性不变的位置上去，现在发展到了所谓斜交解. 这里，不必要求因子相互正交，但要求在其他一些标准下是最优的，如具有显著载荷的变量个数最小. 按照各种不同的旋转标准，对问题的细节作了许多推广，Cattell 是其中的主要论述者之一. 他的工作在本书第 9 章介绍.

*广义方差和协方差分析*是由 Bock 给出的[9]，许多作者进行了实现. 现在需要解决遗失数据中出现的问题. 这里，许多不同的、这样或那样权宜的算法都在应用(如[17]). 目前，对这个问题来讲，还没有一个在应用上是满意的、在理论上是完善的算法. 可以设想，这类算法实际上可能不存在.

b. 计算

自从大存储量的快速数字计算机出现以来，因子分析和多元方差、协方差分析都得到了非常广泛的应用. 对因子分析来讲，计算格式仍在变化，本书第 7 章 Joreskog 的工作就是很好的一例，说明如何从基础数学出发，进行认真细致的考虑和怎样才能得到一个更清晰、更有效的算法.

多元方差和协方差分析已有许多著者进行讨论并编制了程序，其中著名的有 Dean Clyde[10] 和 Jeremy Finn(本书第 10 章). Finn 最新给出的格式允许空格和对照向量的符号表示. 这些所希望的性质，利用别的文本很难实现. 大家知道，多元方差和协方差分析和第二篇里讨论的多元回归分析及多元判别分析有着密切的联系. 事实上，可以把回归分析、判别分析看作方差分析的一个特例. 由于我们希望回归分析程序和判别分析程序都具有一定的灵活性，但这又很难在广义方差分析结构中实现，因此在本书中把它们当作联系不太紧密的两个问题分别进行讨论. 由于这种灵活性至今尚未实现，所以把它作为一个非常困难的计算问题留给读者.

5. 聚类分析和模式识别

a. 理论和算法

聚类分析，特别是谱系聚类，是从 Sokal 和 Sneath 的«数值分类学»[18](1957)一书正式开始的. 早期使用的一些聚类方法仅限于一些常规方法. 近期使用的一些方法，也还是一些类似于数值分类学中的方法，包括最短距离、最长距离以及由树状图(树簇)发展起来的一些类似算法. 在这个领域中，最早出现的问题是如何给不同对象间的对"相似性"有贡献的不同特性加权？它们应该是等权的呢，还是和它们的方差、它们对总均值的贡献等成比例？目前一致认为，除非存在先验信息说明这一变量比其他变量应有更高的权，否则，就应该取成等权的，即所有的权都为 1.

在不同的聚类分析和模式识别中，需要不同的距离函数. 例如，本书中 Lance 和

Williams(第 11 章)介绍的谱系分类法用的是相似系数，基于找出一个子集合使其最接近于一个超球的方法就必须利用欧几里德距离或 Mahalanobis 距离. 当然，我们可以把后面两种测度看成是更一般相似测度的一个子集. 事实上，尽管在逐对距离一经计算好以后，就失去了对变量的识别，但欧几里德距离仍然在许多谱系算法中得到应用. 对有些方法，如本书第 13 章讨论的 ISODATA 方法就不是这样. 在 ISODATA 中，保留了对变量的识别.

象前面提到的那样，可把聚类和模式识别中的方法分成为无师可学和有师可学的两类不同的分类方法. 本书讨论的所有聚类方法都是无师可学型的，换句话说，它们并不要求指定模式或典型观测而想找到基本结构. 在这一考虑中，它们可能同称为 Perceptron[19,20,21] 的更一般的模式识别系统有关. 这一方法作者已经实现[22]，象 ISODATA 那样，看成是一种逐段线性分类的方法.

在谱系分类和 ISODATA 分类中，假若定义了观测的类别，用一些参数，如在判别分析(第 5、6 章)中找到的一组最好的判别函数，分开新出现的各类，这是一种最常用的方法. 在分类过程中，通常会出现一些多元异常点，即一些不能简单拟合其余观测点的数据，这常常意味着存在一些具有很少成员的类.

我们早已提到，特征或者变量具有不同尺度的问题尚未解决. 许多方法，如 ISODATA，对不同的变量暗定有不同的权. 在 ISODATA 方法中，对具有最大正则化标准差的变量给予更大的权. 显然，谱系分类就不是这样，因为在那里已失掉了对变量的识别.

本书第 12 章介绍的多维尺度转换法，通常把它看成是一种模式识别的方法，但事实上，它更接近于寻求有关变量结构的方法. 应该指出，任何一种聚类方法，把数据矩阵旋转 90°，就可以用来寻求观测数据或变量的模式. 当然，在多维尺度转换方法中，也有类似情况存在. 读者无疑会感到奇怪，为什么不把这里讨论的方法归到主成分分析和因子分析中去. 虽然多维尺度转换及与其有关的一些方法更接近于显示结构的非线性方法或象前面提到的类似于非线性因子分析方法，但从实际考虑，我们相信还是应该把它们包括在聚类分析和模式识别的这一部分(即本书的第 IV 部分)里.

b. 计算

现在，有不少实现谱系分类的程序. 很多数据分析学家都以极其相近的程序编写过这类文本. 在这本书里，我们选用了 Lance 和 Williams 的广义谱系分类算法. 使用这种灵活性很强的算法，可以组织这一过程给出或松或紧的、把空间或多或少分开的聚类. 由于 ISODATA 程序是最早把总体分成一些接近于多元正态子集的算法之一，所以本书也选进了这一算法. 本书给出的计算方案和早期实施的方案相比，已经作了一些重大的修改. 虽然在一个混合总体中寻求多元正态子集的问题尚未提出，但 Wolfe[23] 对这个问题已平行地作了不少努力. 进行类似工作的还有 Michigan 大学，定名为 AID[24]. 在某种程度上，这种方法不同于有师可学的回归分析方法. 本书中给出的多维尺度转换过程是由该方法的创始人之一介绍的，代表着文献中给出的很大一部分方法. 为简单起见，建议读者去读本书的第 12 章. 那里，对各种不同的方法有着完整详尽的说明.

6. 时 间 序 列

在时间序列分析中，理论、算法和计算之间密切相关. 所以，在这一节中，把它们放在一块进行讨论.

用自相关和互相关研究时间序列，特别在经济学中研究时间序列，已有多年的历史，但也很奇怪，在生物学中很少应用．更复杂的一些分析，通常包括频率和/或相平面变换，如Fourier 变换，直到最近才在数字计算机上广泛实现，用它们代替模拟计算装置进行滤波和其他一些类似运算．快速 Fourier 变换(FFT)完全改变了这种面貌．有关 FFT 的问题，将在本书第 14 章进行讨论．FFT 的出现，大大降低了原来需要的若干个平方数量级的计算工作量．事实上，运算量的减少是如此之大，以致可以制造各种经济实惠的专用计算装置实现这种变换．因此，能够看到这种奇怪的现象，在通用机和专用机上都能有效地实现这些变换．这里，选择的主要根据是方便．由于 FFT 的出现，混合计算机可能是解决这类计算问题的最好方法．这又回到了原先的计算情况，例如，在模拟装置上进行 Fourier 变换被积函数的处理，在模拟装置的输入端和输出端上装备以数字装置(混合计算机)．

时间序列的预报问题，即从各个变量过去的性态预报它们未来的性态，多年来一直限于一元序列．混合计算机和 FFT 的出现，可以把这些方法推广到多元时间序列[25]．值得注意的是，多元时间序列分析和多元时间序列预报在不同的应用领域里有着不同的名称．例如，在电子工程技术问题中，长期把时间序列分析称为瞬变分析；在环境领域中，称为干扰分析，即分析干扰传播的规律，切断干扰源等问题；在生物学中，“季节”、“周期”、“循环”等术语是常用的；还可以找到其他各种不同的例子．但所有这些方法，实际上处理的是相同的基本问题，即不管时间是在宏尺度上还是在微尺度上进行测量，它们都是研究参数随时间的飘移．

在 Bacon 和 Broekhoven 合编的第 15 章中，对时间序列预报问题有详尽的讨论．

快速数字计算机对实现各种时间序列分析有很大影响．虽然时间序列分析比回归分析和判别分析的发展差不多晚了十年，但和回归分析、判别分析在应用方面的发展是平行的．时间序列分析的真正潜力尚待挖掘，但近些年来发展的步子很大．在今后几年里，时间序列分析会有一些重大的进展．

7．方法的限制及其特点

在表 1 中，粗略地勾划出了本书讨论的各种统计方法的限制及其显著特点．这是一种尝试，正如在分类中进行的尝试一样，距要求相差尚远．所以，表 1 中给出的各项说明，只能作为引玉之砖，而决非定论．

这里，我们对表 1 中“代价”一栏作一些说明．我们想用“代价”这个参数刻划一下影响代价的因素，并根据代价的大小，研究对方法的限制．这里，特别强调的是后者．在逐步回归方法中，没有标出“代价”这一项，意思是说，在回归分析中变量的个数和观测次数都影响代价，但一般说来，它们并不是一种限制．

从表 1 可以看出，在大字长的计算机上没有精度问题．特别，在目前情况下，在 60 位二进制的数字计算机上，对统计计算的精度来讲，根本不存在什么问题．

8．可 用 程 序

在表 2 和表 2a 中，给出本书中一些方法的已有程序的来源．这里，只想给出一些最常用程序的来源，并不打算给出所有可能程序的来源．本书中列出的程序都是用 Fortran 语言编制的，我们并不想以此限定使用计算机的类型或源程序的范围．

表2 可用程序表

章别	3 子集选取	4 逐步回归	5 逐步判别	6 组合方法	7 因子分析 (Joreskog方法)	8 因子分析 (Minres方法)	9 唯一旋转	10 多元方差和协方差分析	11 谱系分类	12 多维尺度转换	13 ISODATA	14 FFT	15 预报
作者	X	X	X	X	X	X	X		X	X	X		X
BMD		X	X										
SPSS		X	X										
Omnitab		X	X										
Universities		X	X					X					
Genesee	X	X	X	X	X	X	X	X	X	X	X		X
Bell Telephone Labs										X			
Cybernet		X	X										
Multiple Access		X	X									X	
IBM		X	X									X	
UNIVAC		X	X										
Xerox		X	X										
Educational Test Serv.					X	X							
CSIRO									X				

表2a 可用程序及其来源

BMD	BMD 程序库 加利福尼亚大学 卫生科学计算研究室	Multiple Access	多伦多数据中心
SPSS	芝加哥大学 国家鉴定研究中心	IBM	IBM T. J. Watson 研究中心
OMNITAB	国家标准局	ETS	普林斯顿 教育检验中心
GENESEE	Genesee 计算中心	P-STAT	普林斯顿大学 普林斯顿计算中心
Bell	贝尔电话实验室	CSIRO	澳大利亚 畜牧研究实验室
Cybernet	中西部控制中心	XEROX	Xerox 有限公司

9. 参考文献

[1] C. R. Rao, "Recent Trends of Research Work in Multivariate Analysis," Biometrics, **28**, 3～22 (1972).

[2] K. Enslein, "Computer-Based Data Analysis in Longevity Research," *Comput. Biomed. Res.*, **3**, 289～329 (1970).

[3] K. Enslein, "Adjustment for Physician Variability and Placebo in Drug Evaluation"(in press).

[4] H. H. Harman, *Modern Factor Analysis*, Univ. of Chicago Press, 1960 and 1971.

[5] R. A. Fisher, *Statistical Methods for Research Workers*, Hafner Publishing Co., New York, 1958.

[6] M. A. Efroymson, "Multiple Regression Analysis," in *Mathematical Methods for Digital Computers*, Vol. 1, A. Ralston and H. Wilf (eds.) Wiley, New York, 1964.

[7] N. Mantel, "Why Stepdown Procedures vs. variable Selection," *Technometrics*, **12**, 621～625 (1970). "More on Variable Selection and an Alternative Approach," **13**, 455～457 (1971).

[8] A. E. Hoerl and R. W. Kennard, "Ridge Regression: Biased Estimation for Nonorthogonal Problems," *Technometrics*, **12**, 55～67 (1970).

[9] R. D. Bock, "Programming Univariate and Multivariate Analysis of Variance," *Technometrics*, **5**, 95～117 (1963).

[10] D. J. Clyde, E. M. Crammer, R. J. Sherin, *Multivariate Statistical Programs*, Biometric Lab., University of Miami, Coral Gables, Fla. 33124, 1966.

[11] J. Arvesen and D. Salsburg, *Approximate Tests and Confidence Intervals using the Jackknife*, Mimeograph series 267, Purdue University Statistics Dept., 1971.

[12] D. Gray and Schucany, *The Generalized Jackknife Statistic*, Marcel Dekker, New York, 1972.
[13] F. Mosteller, "The Jackknife,"*Rev., Inter. Stat. Institude*, **39**, 363(1971).
[14] W. J. Dixon (Ed.), *BMD Biomedical Computer Programs*, University of California Press, Berkeley, California, 1960 and 1970.
[15] Genesee Computer Center, Rochester, New York, 1973.
[16] H. Hotelling, "Analysis of a Complex of Statistical Variable into Principal Components," *J. Educat. Psychol.*, **26**, 139～142(1935).
[17] K. Enslein, *A Novel Method for the Resolution of the Missing Observation Problem*, Spring Meeting of ENAR, Chapel Hill, N. C., 1969.
[18] R. R. Sokal and P. A. H. Speath, *Principles of Numerical Taxonomy*, W. H. Freeman and Co., San Francisco, Cal., 1963 and 1973.
[19] F. Rosenblatt, *Principles of Neurodynamics*, Spartan Books, Washington, D. C., 1962.
[20] H. D. Block, B. W. Knight Jr., and F. Rosenblatt, "Analysis of A Four-Layer Series-Coupled Perceptron, II," *Rev. Mod. Phys.*, **34**, 135～143(1962).
[21] M. Minsky and S. Papert, Perceptron, *An Introduction to Computational Geometry*, MIT-The Science Press, Cambridge, Mass., 1969.
[22] K. Enslein, "A General-Purpose Perceptron Simulator," *Comput. Biomed. Res.*, **1**, 187～214(1967).
[23] J. H. Wolfe, "Patter Clustering by Multivariate Mixture Analysis," *Multivariste Behav. Res.*, **5**, 329～350(1970).
[24] J. A. Sonquist, *Searching for Structures (Alias Aid-III)*, Institute for Social Research University of Michigan, Ann Arbor, Michigan, 1971.
[25] J. S. Bendat, A. G. Piersol, *Measurement and Analysis of Random Data*, Wiley, New York, 1966.