

多元统计分析方法在林业中应用及 IBM-PC系列程序集编制的研究*

李希菲

(中国林业科学研究院资源信息研究所)

关键词 多元统计分析; 矩阵运算; 软件包

从70年代开始,随着微机的普及、推广,多元统计分析方法及其他数值分析方法在各个领域得到越来越广泛的应用。这就产生了两种要求:一是系统总结多元统计分析方法理论和应用情况,评价它在林业中应用的可能及前途;二是要为微机配置一套通用程序,使林业工作者从繁重的计算工作中解放出来,使多元统计分析方法付诸实施,也有利于多元统计分析方法及其他数值分析方法进一步推广和应用。本课题的研究从1982年开始,至1988年完成。今年元月通过中国林科院院级鉴定,并获得有关专家的好评。

一、理论部分——《多元统计分析方法》

我国70年代末引进了多元统计分析方法,70年代末、80年代初在工业、医学、地质等方面得到应用。在林业中除在遥感及生态方面有少量应用外,其他领域尚不多见。《多元统计分析方法》系统介绍了多元统计分析这个学科和它的主要内容、计算方法及应用概况,分析了多元统计分析方法在林业中应用的可能性,并附有实例,方便初学者试用。同时,作者还对一些多元统计的基础理论问题进行了研究,取得一些进展。例如关于主坐标分析的基础理论,主坐标分析与距离的关系,数量化方法与数据结构及与其他统计方法的联系等内容。近年来,林学界已普遍将多元统计分析方法作为科研、生产设计、数据分析的重要手段之一,许多高等林业院校已将它列入有关专业研究生的必修课之一。

二、矩阵运算器 OPMA

矩阵运算器是处理矩阵数据的强有力工具,采用命令集方式编程,包含18个命令、30个函数,可以直接用矩阵公式进行矩阵运算。特别适用于学校作为教学及学生实习用。它与程序集(见下节)提供的数据变换程序相结合,为数据处理提供了很方便的工具。

本文于1990年2月9日收到。

*本课题由唐守正负责。参加人有郎奎健(东北林业大学)、刘继红、李希菲、洪玲霞、张淑娟、蔡力平(东北林业大学)等同志。本文由李希菲执笔。

三、应用程序集

程序集有二种编制方法:一种是命令集式,例如欧美的SAS、ANALYST、BMOP、NCSS等,这类程序结合性能较好,但使用较复杂。另一种是人机对话式程序,比较适合对计算机不太熟悉或初学者使用。但这种程序没有组合联接功能,我们的程序集采用了人机对话的编程方式,编制了90个程序。它包括了较全面的多元统计分析方法程序,例如多元样本的假设检查、多变量回归及逐步回归、聚类分析、判别分析、主成分分析、主坐标分析、典型分析及各种数量化方法。除此之外,还包括线性规划、动态规划及经营管理的一些主要方法,例如统计及矩阵运算的主要内容,均可由本系统处理。数据变换程序有较强的数据变换功能:可以追加、合并、变换、条件处理已有的数据,为各程序使用同一组数据提供了有力手段。

(一) 程序集的特点

1. 通用性 在程序集里,通用曲面回归是一个功能很强的通用程序,它提供了3种估计任意连续曲线参数值的方法。其中改进平行线法甚至可以计算样条回归以及不可微连续曲线的回归参数估计问题。现在世界上只有少数几个为大型机配置的大规模程序包,如SAS才具有这种功能,但仍然不能计算不可微曲线的回归问题。在微机上具有这种功能尚属少见。

广义方差分析程序几乎可以处理单因变量的所有参数统计问题,包括回归分析、数量化方法I、平均数的假设检查、各种分组方法的方差分析、协方差分析、各种正交及非正交实验设计的分析、区组实验、拉丁方实验等实验数据的分析。

2. 专用性 三次样条标准木法求树干材积是近代测树学的一项新技术,树干解析I型、II型、材积表法求生长量等是测树专业常用的方法。立地指数I型、II型可以对经常使用的各种线型的导向曲线同时算出立地指数族,在本专业内具有通用性。

(二) 本程序集与同类程序集比较

表1列出了本程序集与国内外几个以统计为主的程序包的功能比较。仅统计部分而言,从表中可见优于国内一些地方的程序集。与国外微机上的程序集相比各有千秋,功能接近中小型机上程序集,但曲面(包括曲线)回归程序是本程序集特有的。此外还包括有关林业、规划及管理方面的内容。

(三) 程序研制过程中的理论工作

在程序研制过程中,解决了计算方法及设计原理方面的两个理论问题:

1. 改进的平行线法计算曲面回归参数,这一方法是对华罗庚教授倡导的0.618法及平行线法(二维0.618法)向多元回归问题的推广,并且在程序设计中加进了区间函数 $FNT(A, B, X)$,使这个方法可用于样条回归及不可微函数的回归问题,通用曲面回归之所以功能超过其他软件包,主要依赖于这一理论问题的解决。

2. 广义方差分析,是把1973年Rao提出的一般线性模型理论具体用于方差分析及协方差分析,因而解决了不平衡数据、样本丢失等资料的方差分析及协方差分析问题。这一理论与计算机编程技术相结合,使广义方差分析程序成为一个非常通用的程序。

(四) 应用效果及今后设想

《多元统计分析方法》出版后,书中介绍的方法已被广泛引用,但多元统计分析方法所涉

表 1 国内外部分统计程序包功能比较

程 序 名	ANALYST	NCSS	RDAS	IBM PC系列程序集
研制单位	富士通	J. L Hintze	上海市计算所	中国林科院资源信息所
适用机型 输入方式	大中型 终端、文件	微机 终端、文件	小型机、微机 终端、文件	微机 终端、文件
数据变换 存储编辑	四则、函数、IF语 句、样本删除、追加 样本、合并文件	四则、函数、IF语 句、追加样本	运算、计数、IF语 句	四则、函数、IF语句、变 量增减、变换、追加样本、 样本删除、合并文件
基本矩阵运算	不可行	不可行	不可行	可 行
图 形	较强散点图、直方 图、无曲线图	无	无	散点图、曲线图、部分 直方图
分组统计	有	有	无	有
线性回归	完善、无双重筛选	完善、无双重筛选	完善、无双重筛选	完善、有双重筛选
非线性回归	无	无	无	有
一般数理统计案 类、判别	有	有	有	有
典型分析	有	无	无	有
因子分析	有	有	有	无
数量化理论	有	无	无	有
时间序列	无	部 分	部 分	无
非平衡数据分析	无	无	有	有
线性规划和动态规 划	无	无	无	有

及的基础理论较深，还需进一步普及和推广，才能不断提高我们的科研和生产水平。

程序集推出已5年，全国有60多个正式用户，确实给一些科研工作者和课题组解决了数据分析问题，提高了科研水平。就我们森林经理和林业统计研究室来看，程序包对我们几年来科研工作的顺利完成起了很大作用。例如“全国用材林发展趋势的研究”这一课题，共配置了全国40多个树种(组)的生长过程，如用手工计算大约30人要半年才能完成，而用程序包2个人半个月就完成了。在一些生长模型的建立和一些数表的编制方面，如编制可变密度收获表，要建立林分断面面积与立地、年龄、密度等因子的相关模型，通常情况下很难解决，采用通用曲面回归可以完成这一模拟，使我们实现了利用先进技术编表和编制先进的生产用表。

但是，比较高级的程序要求使用者有较高的水平，因此应当不断提高科研、生产人员的素质，大力推广和普及多元统计分析方法，对程序用户举办各种类型的学习班，并对用户进行跟踪调查，才能不断提高我们的科研和生产水平，不断完善我们的程序功能。

参 考 文 献

- [1] 唐守正, 1986, 多元统计分析方法, 中国林业出版社。
[2] 郎奎健等, 1989, IBM-PC 系列程序集, 中国林业出版社。

APPLICATIONS OF MULTIVARIATE STATISTICAL ANALYSIS IN FORESTRY AND ITS IBM-PC SOFTWARE PACKAGES

Li Xifei

(The Research Institute of Forest Resource Information Techniques CAF)

Abstract A project of applications of multivariate statistical analysis in forestry and its software packages has been engaged for six years. Three achievements, a book and two software packages have been submitted.

1. The book "Methods of multivariate statistical analysis" was written to provide researchers, investigators and workers in forestry and agriculture with an elementary and important multivariate technique. The book includes many examples which have been employed in forestry and agriculture successfully.

2. The IBM-PC program set consists of 90 programs, which involve statistics, programming and management. The set not only includes usual data transformations, statistical and multivariate analysis programs but also has some powerful programs, e. g. general multivariate nonlinear regression and general ANOVA. All the programs are easy-to-use and source programs are provided.

3. OPMA (MAtrix Operator) is a software package for dealing with matrix operation. OPMA makes matrix operation easy. It is a useful tool to compute matrix and suitable for the students to learn matrix operation and computing methods of multivariate analysis.

Key words multivariate statistics; matrix operation; software packages