



王斌会 著

# 经济管理模型的 多变量统计方法 及分析系统 Qstat

 中国统计出版社  
China Statistics Press

## (京)新登字 041 号

### 图书在版编目(CIP)数据

经济管理模型的多变量统计方法及分析系统 Qstat/王斌会著.  
- 北京:中国统计出版社,2005.6

ISBN 7-5037-4626-2

I. 经…

II. 王…

III. ①经济管理-经济模型-多变量-统计方法-教材

②经济统计-统计分析-应用软件, Qstat-教材

IV. F222.1

中国版本图书馆 CIP 数据核字(2005)第 050147 号

经济管理模型的多变量统计方法及分析系统 Qstat

---

作 者/王斌会

责任编辑/吕 军

装帧设计/艺编广告·赵 辉

出版发行/中国统计出版社

通信地址/北京市西城区月坛南街 75 号 邮政编码/100826

办公地址/北京市丰台区西三环南路甲 6 号

电 话/(010)63459084 63266600-22500(发行部)

印 刷/河北天普润印刷厂

经 销/新华书店

开 本/850×1168mm 1/32

字 数/318 千字

印 张/13.375

印 数/1-2000 册

版 别/2005 年 8 月第 1 版

版 次/2005 年 8 月第 1 次印刷

书 号/ISBN 7-5037-4626-2/F·2024

定 价/27.00 元

---

中国统计版图书,版权所有。侵权必究。

中国统计版图书,如有印装错误,本社发行部负责调换。

# 前 言

多变量统计分析简称多元统计分析,是数理统计学的一个重要分支,具有很强的应用性,是进行科学研究的一项重要工具,在自然科学、社会科学和经济管理等领域中得到越来越广泛的应用。随着计算机的普及和统计软件的广泛使用,了解和运用它的人迅速增加。作为一种数据处理的非常有用的方法,在各个领域都取得了卓有成效的成果。

多变量统计分析可以看做是处理多个变量的综合统计分析方法,它可以把多个变量对一个或多个变量的作用程度大小线性地表示出来,反映事物多变量间的相互关系;可以消除多个变量的共线性,将高维空间的问题降至低维空间中,在尽量保存原始信息量的前提下,消除重叠信息,简化变量间的关系;可以通过事物的表象,挖掘事物深层次的、不可直接观测到的属性即引起事物变化的本质;也可以透过繁杂事物的某些性质,对事物进行识别、归类。

国内目前出版的多变量统计分析专著不是很多,适合于经济

管理类学生使用的教材也较少。本书的编写目的是提供一本适合于财经管类院校本科生和研究生使用的参考书、教材和软件使用手册。

众所周知,多变量统计分析是以概率统计这门数学科学为基础,应用线性代数的基本原理和方法并结合计算机对实际资料和信息进行收集、整理和分析的一门科学。因此,它的原理较为抽象,对学生的数学基础要求也较高,教学中存在着大量的数学公式、数学符号、矩阵运算和统计计算,必须借助于现代化的计算工具。

在目前的统计学教学、科研和应用中,我们深切感到教学与实践是脱离的,教材中的内容和统计软件中的内容在设置上不完全一致,对数据格式的安排也不仅相同,对资料的要求也有一些差别。研究人员大都只需了解这些统计方法的使用,并不需要掌握其理论推导。目前多媒体教学课件的设计仍以电教人员为主,统计软件设计是以计算机专家为主,都未能充分结合教与学的问题,使得学生或科研人员学完统计学后仍不能使用统计软件,统计软件中的许多方法学生或科研人员又掌握不够,我们认为,统计软件应成为数据处理与教学科研相结合的产物,使数据处理与统计设计相结合,并具有一定的智能化,让数据满足方法,方法符合数据,而不是不考虑条件地瞎用。因此,我们在编写本书的过程中在这方面进行了有益的尝试,力争使多变量统计分析这门有用而复杂的方法变得通俗易懂、操作方便、简单实用。

本书的特色是:

(1)原理、方法、算法和实例分析相结合:鉴于目前计算机统计分析软件已是多变量统计分析应用中不可缺少的工具,本书特别注意各种多变量统计算法的实现,使得给出的算法更有实用的价值。为此,在论述算法思想时就引进易于化为计算步骤的数学式子和符号,并在计算步骤中采用了我们开发的统计软件。

(2)每一章后面都有用我们开发的统计软件进行的综合案例分析:本书在讲清各种方法的实际背景和数学思想的同时,对每种方法都给出具体的经济管理实例,并结合统计软件进行案例分析。书中的大多数案例都是作者收集的最新实际数据。

(3)解决统计软件用于统计学教学和科研中存在的问题:国内目前缺乏适合开展多变量统计分析教学科研的统计分析软件,SAS、SPSS 等国外统计软件,一是没有版权,需要昂贵的费用购买,二是全为英文软件,使用复杂,与教课书内容设置不完全一致,财经管类学生和研究人员使用尤其困难。于是我们开发了一套用于多变量数据分析的统计软件解决了这个问题。

### 本书的创新点:

(1)提供了一套用于多变量统计分析的计算机软件:作者通过近6年的努力,自主开发了一套具有知识产权的统计分析软件 Qstat,本书的所有方法都是用其实现的,所有实例及案例都是用其分析的。书中的所有结果、图形都是由 Qstat 给出的。高质量的图形和方便的结果输出是其他统计软件所不能比拟的。

(2)研究如何将统计软件的数据处理与统计教学相结合,形成一套完整的教学科研相结合的初具智能化的多媒体统计软件。在教学与科研一体化的功能上,在数据编辑、统计分析、统计设计、统计绘图和统计帮助上充分体现多媒体教学的特点。所以本书也可以用于多媒体教学教材。

(3)在统计分析方面,常用的软件有 SAS、SPSS 和 Minitab 等。本书提供的由我们开发的有完全自主产权的统计分析软件 Qstat,在应用上,具有其他软件不可比拟的操作简单、接口方便、扩充能力强等优势。例如在聚类图、判别图、因子图 and 对应图的绘制,综合评分的计算等方面都有许多新的特点。Qstat 实际上可以应用到统计学的许多领域。多变量统计分析虽然复杂,但 Qstat 平台将引导读者走向成功!

本书的内容安排吸收了国内外有关多变量统计分析论著的特点,在章节的安排上遵循由浅入深、由简到繁的原则,对多变量统计量和分布进行了较为详细的介绍,增加了一些多变量统计推断的内容,同时附加了一些线性代数和矩阵运算的概念。书中的主要内容是根据我在暨南大学多年从事多变量统计分析课程教学所撰写的讲稿的基础上修订而成的,书中一些内容也是作者多年从事多变量统计分析的研究成果,例如关于变量选择准则的研究、稳健回归算法的研究,岭回归算法的研究,多重共线性诊断标准的探讨等。许多新理论和方法还在不断完善中,限于时间将在以后的版本中陆续加入。本书软件的计算结果大都经过统计分析软件 SAS 和 SPSS 考证,结果可靠。

本书由作者一人承担完成,书的内容部分得到广东省科技计划项目《经济管理模型的算法研究及预测决策支持系统研制:2004B10101010》的资助,统计系研究生陈一非、陈楚祥、吕亮雯、毛艺萍等同学为本书的撰写提供了一些帮助,陈萍、陈光慧、孔丽娜、刘畅、谢铭杰、汤克明等同学提供了一些案例数据,在此一并表示感谢。

本书是国内第一个用自己开发软件编写的多变量统计分析论著,由于作者知识和水平有限,书中难免有错误和不足之处,恳请读者批评指正!

王斌会

2004 年 11 月于暨南花园

# 目 录

## 第 1 章 多变量统计分析简介/1

§ 1.1 多变量统计分析的来历/1

§ 1.2 多变量统计分析的内容/2

§ 1.3 多变量统计分析的用途/7

§ 1.4 如何收集和整理多变量分析资料/8

练习题 1/12

## 第 2 章 随机向量及其分布/13

§ 2.1 随机变量及其分布回顾/13

§ 2.2 随机向量及其分布/16

2.2.1 随机向量的期望和方差/16

2.2.2 随机向量二次型的均值和方差/17

2.2.3 随机向量的独立性和卡方分布/19

§ 2.3 多变量正态分布/20

- 2.3.1 多变量正态分布的定义/21
- 2.3.2 多变量正态随机变量的母函数/23
- 2.3.3 正态随机变量的二次型/25

练习题 2/28

### 第3章 多变量数据的直观分析法/29

- § 3.1 直条图/30
- § 3.2 箱尾图/31
- § 3.3 轮廓图/32
- § 3.4 雷达图/33
- § 3.5 星相图/35
- § 3.6 调和曲线图/35
- § 3.7 星座图/38
- § 3.8 矩阵散点图 /40

综合案例及 Qstat 软件分析/41

练习题 3/44

### 第4章 多变量资料的统计检验/45

- § 4.1 多变量样本均值和协方差的分布/45
- § 4.2 多变量资料的假设检验/51
  - 4.2.1 多变量正态总体分布的检验/52
  - 4.2.2 多变量正态总体均值向量的检验/56
  - 4.2.3 多变量正态总体协方差阵的检验/67

综合案例及 Qstat 软件分析/68

练习题 4/73

### 第5章 多变量回归与相关分析/77

- § 5.1 多变量之间关系分析/77
- § 5.2 多变量回归分析/78



5.2.1 单变量回归分析回顾/78

5.2.2 多变量回归模型的建立/85

§ 5.3 多变量相关分析/93

5.3.1 简单相关分析/93

5.3.2 偏相关分析/98

5.3.3 复相关分析/102

综合案例及 Qstat 软件分析/103

练习题 5/110

## 第 6 章 多变量回归分析的推广应用/115

§ 6.1 回归变量的选择方法/115

6.1.1 回归变量的选择准则/115

6.1.2 逐步回归分析方法/121

§ 6.2 回归诊断方法简介/129

6.2.1 误差的正态性检验/129

6.2.2 异常值的产生原因、检测及处理/130

6.2.3 多重共线性的检测及处理/136

§ 6.3 抗异常值的稳健 M 估计方法/140

§ 6.4 抗多重共线性的岭回归方法/142

综合案例及 Qstat 软件分析/146

练习题 6/158

## 第 7 章 聚类分析/160

§ 7.1 聚类分析的概念/160

§ 7.2 聚类分析的类型/161

§ 7.3 聚类统计量/162

§ 7.4 系统聚类法/166

7.4.1 系统聚类法的基本思想/166

7.4.2 系统聚类法的计算公式/167

7.4.3 系统聚类法的基本步骤/169

7.4.4 聚类分析的一些问题/171

综合案例及 Qstat 软件分析/177

练习题 7/183

## 第 8 章 判别分析/188

§ 8.1 判别分析的概念/188

§ 8.2 距离判别/189

8.2.1 两总体距离判别/189

8.2.2 多总体距离判别/191

§ 8.3 Fisher 线性判别/199

§ 8.4 典型判别分析/204

§ 8.5 Bayes 判别法/205

8.5.1 Bayes 判别准则/205

8.5.2 正态总体的 Bayes 判别/207

综合案例及 Qstat 软件分析/212

练习题 8/227

## 第 9 章 主成分分析/231

§ 9.1 主成分分析的直观解释/232

§ 9.2 主成分分析的性质/234

§ 9.3 主成分分析的步骤/238

§ 9.4 应用主成分注意事项/245

综合案例及 Qstat 软件分析/246

练习题 9/255

## 第 10 章 因子分析/262

§ 10.1 因子分析的思想/262

§ 10.2 因子分析模型/265

- § 10.3 因子旋转/273
- § 10.4 因子得分/278
- § 10.5 因子分析的步骤/281
- § 10.6 实际中如何进行因子分析/290
- 综合案例及 Qstat 软件分析/294
- 练习题 10/303

## 第 11 章 对应分析/309

- § 11.1 对应分析的提出/309
- § 11.2 对应分析的基本原理/310
- § 11.3 对应分析的计算步骤/313
- § 11.4 运用对应分析应注意的几个问题/326
- 综合案例及 Qstat 软件分析/327
- 练习题 11/337

## 第 12 章 典型相关分析/340

- § 12.1 引言/340
- § 12.2 典型相关分析的基本构架/341
- § 12.3 典型相关分析的基本原理/342
- § 12.4 典型相关系数及变量的计算/346
- 综合案例及 Qstat 软件分析/350
- 练习题 12/358

## 第 13 章 非线性回归分析/362

- § 13.1 单变量非线性回归模型/362
  - 13.1.1 曲线回归 - 可化为线性回归的非线性回归/362
  - 13.1.2 曲性回归模型的具体应用/367
- § 13.2 多变量非线性回归模型/376
  - 13.2.1 非线性回归模型概述/376

|                             |
|-----------------------------|
| 13.2.2 非线性回归模型最小二乘估计的计算/377 |
| 综合案例及 Qstat 软件分析/383        |
| 练习题 13/387                  |

## **附录 1 矩阵代数的基本知识/391**

- § 1. 矩阵的定义和代数运算/391
- § 2. 矩阵代数运算的性质/392
- § 3. 矩阵的转置和对称矩阵/393
- § 4. 逆矩阵与正交阵/394
- § 5. 矩阵的初等行变换和初等矩阵/395
- § 6. 矩阵的秩/397
- § 7. 特征根和特征向量/398
- § 8. 正定阵、非负定阵和投影阵/400
- § 9. 矩阵的微商/402

## **附录 2 统计分析软件 Qstat 使用说明/404**

## **参考文献/412**

# 第 1 章 多变量统计分析方法简介

## § 1.1 多变量统计分析的来历

在统计学的基本内容中,只考虑了一个因素或几个因素对一个观测指标(变量)影响大小的问题,称为单变量统计分析或单因素分析。若考虑一个因素或几个因素对两个或两个以上观测指标(变量)的影响大小,或者多个观测指标(变量)间的相互关系问题,称为多变量统计分析。多变量分析是研究客观事物中多个指标(变量)间相互依赖的统计规律的一个数理统计学分支。

在经济生活中,受多种指标(随机变量)共同作用和影响的现象大量存在。有两种方法可同时对多个随机变量的观测数据进行有效地分析和研究。一种做法是把多个随机变量分开分析,每次处理一个去逐次分析研究。但当变量较多时,变量之间不可避免地存在着相关性,如果分开处理不仅会丢失很多信息,往往也不容易取得好的研究结论。另一种做法是同时进行分析研究,即用多变量统计分析方法来解决,通过对多个随机变量观测数据的分析,来研究变量之间的相互关系以及揭示变量的内在规律。所以说,多变量统计分析就是研究多个随机变量之间相互依赖关系以及内在统计规律的一门学科。

多变量统计分析是运用数理统计方法来研究解决多指标问题的理论和方法。构成多变量统计分析模型的数学方法并不新颖,

如与多变量有关的基本概率分布(多变量正态分布就源自 30 年代),主成分分析是首先由 K.Pearson 于 1901 年提出,再由 Hotelling(1933)加以发展的一种统计方法。由于当随机变量较多时,多变量分析的计算工作量极端繁冗,没有计算机根本无法完成。因此,直到有了计算机之后,多变量分析技术才进入实用阶段并迅速发展。近 20 年来,随着计算机应用技术的发展和科研生产的迫切需要,多变量统计分析技术被广泛地应用于经济、管理、地质、气象、水文、医学、工业、农业和教育学等许多领域,已经成为解决实际问题的有效方法。

## § 1.2 多变量统计分析的内容

多变量统计分析包括的主要内容有:多变量的数字特征及其分布;多变量总体的参数估计和假设检验;多变量数据图示法;多变量线性相关与回归分析;聚类分析;判别分析;主成分分析;因子分析;对应分析;典型相关分析等。

**1. 随机向量及其分布:**很多随机现象中,对一个随机试验需要同时考察几个随机变量,例如发射一枚炮弹,需要同时研究弹着点的几个坐标;研究市场供给模型时,需要同时考虑商品供给量、消费者收入和市场价格等因素。

**2. 多变量总体的参数估计与假设检验:**考察某商业行业今年和去年的经营状况,这时需要看这两年经营指标的平均水平是否有显著差异就需要用多变量均值向量的  $T^2$  检验。如果要考虑几年的经营水平之间的差别,就需用多变量方差分析。

**3. 多变量的直观分析:**多变量直观分析即图示法是进行数据分析的重要辅助手段。例如通过两变量的散点图可以考察异常的观察值对样本相关系数的影响,利用矩阵散点图考察多变量之间的关系,利用多变量箱尾图可以比较几个变量的基本统计量大小差别。

**4. 多变量相关分析:**相关分析就是要通过对大量数字资料的

观察,消除偶然因素的影响,探求现象之间相关关系的密切程度和表现形式。在经济系统中,各个经济变量常常存在密切的关系。例如,经济增长与财政收入,人均收入与消费支出等。在这些关系中,有一些是严格的函数关系,这类关系可以用数学表达式表示出来。还有一些是非确定的关系,一个变量产生变动会影响其他变量,使其产生变化。其变化具有随机的特性,但是仍然遵循一定的规律。对于函数关系可以很容易地解决,而对那些非确定的关系,称为相关关系,才是我们所关心的问题。

**5. 多变量回归分析:**回归分析研究的主要对象是客观事物变量间的统计关系。它是建立在对客观事物进行大量实验和观察的基础上,用来寻找隐藏在看起来不确定的现象中的统计规律的统计方法。回归分析不仅要揭示因变量对自变量的影响大小,还可以由回归方程进行预测和控制。回归分析的研究范围如下:

- |   |           |          |              |
|---|-----------|----------|--------------|
| { | ①线性回归模型   | {        | 单变量线性回归      |
|   |           |          | 多变量线性回归      |
|   | ②回归模型的诊断  | {        | 回归模型基本假设的合理性 |
|   |           |          | 回归方程拟合效果的判定  |
|   |           |          | 选择回归函数的形式    |
|   | ③回归变量选择方法 | {        | 自变量选择的准则     |
|   |           | 逐步回归分析方法 |              |
| { | ④含定性变量的回归 | {        | 自变量含定性变量     |
|   |           |          | 因变量含定性变量     |
| { | ⑤估计方法改进   | {        | 岭回归          |
|   |           |          | 主成分回归        |
|   |           |          | 偏最小二乘回归      |
|   |           |          | 稳健回归分析       |
| { | ⑥非线性回归    | {        | 单变量非线性回归     |
|   |           |          | 多变量非线性回归     |

**6. 聚类分析:**聚类分析是研究“物以类聚”的一种现代统计分析方法,在社会、人口、经济、管理、气象、地质、考古等众多的研究领域,都需要采用聚类分析作分类研究。例如不同地区城镇居民收入和消费状况的分类研究;区域经济及社会发展水平的分析及全国区域经济综合评价;在儿童生长发育研究中,把以形态学为主的指标归于一类,以机能为主的指标归于另一类。过去人们主要靠经验和专业知识,作定性分类处理,很少利用数学方法,致使许多分类带有主观性和任意性,不能很好地揭示客观事物内在的本质差别和联系,特别是对于多因素、多变量的分类问题,定性分类更难以实现准确分类。为了克服定性分类的不足,多变量统计分析逐渐的被引进数值分类学,形成了聚类分析这个分支。聚类分析也是一种分类技术,与多变量分析的其他方法相比,该方法较为粗糙,理论上还不完善,但应用方面取得了很大成功。与回归分析、判别分析一起被称为多变量分析的三大方法。

**7. 判别分析:**判别分析是多变量统计分析中用于判别样本所属类型的一种统计分析方法。所谓判别分析法,是在已知的分类之下,一旦遇到有新的样品时,可以利用此法选定一判别标准,以判定该新样品放置于哪个类中。判别分析的目的是对已知分类的数据建立由数值指标构成的分类规则,然后把这样的规则应用到未知分类的样品中去分类。例如,我们有了患胃炎的病人和健康人的一些化验指标,就可以从这些化验指标中发现两类人的区别。把这种区别表示为一个判别公式,然后对怀疑患胃炎的人就可以根据其化验指标用判别公式诊断。

**8. 主成分分析:**在实际问题中,研究多变量问题是经常遇到的,然而在多数情况下,不同变量之间是有一定相关性,必然增加了分析问题的复杂性。主成分分析就是一种通过降维技术把多个指标约化为少数几个综合指标的统计分析方法。例如在经济管理中用主成分分析将一些复杂的数据合成几个商业指数形式,如物价指数、生活费用指数、商业活动指数等。又如对全国 31 个省



市自治区经济发展作综合评价,这时显然需要选取很多指标,如何将这些具有错综复杂关系的指标综合成几个较少的成分,既有利于对问题进行分析和解释,又能便于抓住主要矛盾做出科学的评价,就可以用主成分分析方法。

**9. 因子分析:**因子分析是主成分分析的推广,它也是一种把多个变量化为少数几个综合变量的多元分析方法,但其目的是用有限个不可观测的隐变量来解释原始变量之间的相关关系。主成分分析通过线性组合将原变量综合成几个主成分,用较少的综合指标来代替原来较多的指标(变量)。但在多变量分析中,变量间往往存在相关性,是什么原因使变量间有关联呢?是否存在不能直接观测到的、但影响可观测变量变化的公共因子?因子分析就是寻找这些公共因子的统计分析方法,它是在主成分的基础上构筑若干意义较为明确的公因子,以它们为框架分解原变量,以此考察原变量间的联系与区别。例如在研究糕点行业的物价变动中,糕点行业品种繁多,多到几百种甚至上千种,但无论哪种样式的糕点,用料不外乎面粉、食油、糖等主要原料。那么面粉、食油、糖就是众多糕点的公共因子,各种糕点的物价变动与面粉、食油、糖的物价变动密切相关,要了解或控制糕点行业的物价变动只要抓住面粉、食油和糖的价格即可。

**10. 对应分析:**对应分析又称为相应分析,于1970年由法国统计学家 J.P. Beozecri 提出。对应分析是在因子分析基础之上发展起来的一种多元统计方法,是 Q 型和 R 型因子分析的联合应用。在经济管理数据的统计分析中,经常要处理三种关系:即样品之间的关系(Q 型关系)、变量间的关系(R 型关系)以及样品与变量之间的关系(对应型关系)。如对某一行业所属的企业进行经济效益评价时,不仅要研究经济效益指标间的关系,还要将企业按经济效益的好坏进行分类,研究哪些企业与哪些经济效益指标的关系更密切一些,为决策部门正确指导企业的生产经营活动提供更多的信息。这就需要有一种统计方法,将企业(样品)和指标(变