



现代统计分析方法及应用系列丛书

傅德印 张旭东 编著

EXCEL 与多元统计分析

——附实用计算机程序

中国统计出版社
China Statistics Press

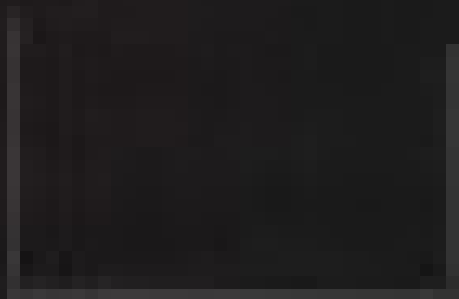


清华大学出版社

EXCEL 与多元统计分析

——附实用计算机程序

清华大学出版社



现代统计分析方法及应用系列丛书

0212.4

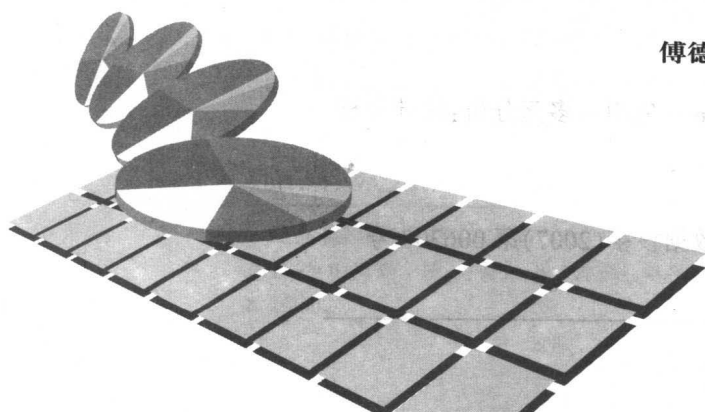
21D

2007

EXCEL 与多元统计分析

——附实用计算机程序

傅德印 张旭东 编著



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

Excel 与多元统计分析/傅德印,张旭东编著.

- 北京:中国统计出版社,2007.3

ISBN 978-7-5037-4289-7

I. E...

II. ①傅...②张...

III. 电子表格系统, Excel - 应用 - 多元分析:统计分析

IV. 0212.4 - 39

中国版本图书馆 CIP 数据核字(2007)第 006307 号

作 者/傅德印 张旭东

责任编辑/王利群 蒋 聪

装帧设计/艺编广告

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 57 号 邮政编码/100826

办公地址/北京市丰台区西三环南路甲 6 号

网 址/www.stats.gov.cn/tjshujia

电 话/邮购(010)63376907 书店(010)68783172

印 刷/河北天普润印刷厂

经 销/新华书店

开 本/787×1092mm 1/18

字 数/440 千字

印 张/23

印 数/1 - 3000 册

版 别/2007 年 5 月第 1 版

版 次/2007 年 5 月第 1 次印刷

书 号/ISBN 978-7-5037-4289-7/O·60

定 价/39.00 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

多年来,我们一直在西北地区从事统计教学工作,深感推动统计学在西北地区快速发展的责任。这份责任不仅在于要推动统计学学科专业自身的建设和发展,而且还在于推动统计理论和方法在非统计专业领域的广泛应用。多元统计分析作为研究多维随机变量之间相互依赖关系以及内在规律的一门科学,它在自然科学、社会科学、人文科学等各个领域都有着广泛的运用,并取得了许多卓有成效的应用成果。多元统计分析已经越来越成为各门学科进行科学研究、数据分析、数据处理必不可少的重要方法,已经成为统计专业教学、统计知识向其它领域进行推广应用的重要组成部分。然而,在我们的教学中,也常常遇到如下的问题:

一是,对于那些旨在驾驭统计分析方法在自己的研究领域中正确运用,而不太关心统计方法如何创新如何发展的人员来说,如何通俗易懂地快速地接受多元统计分析方法的基本思想、方法原理、应用步骤,且不至于用错。

二是,对于那些已经接受并开始运用多元统计分析方法来研究解决本领域具体问题的研究者而言,如何很方便地在计算机上实现各种统计分析方法,特别是对于非统计专业,电脑中没有安装专业统计分析软件的人员,以及那些即使是专业统计人员但由于资金限制而无法使用昂贵的正版专业统计软件时,如何方便、快速、节约地在计算机上实现多元统计分析,无疑是十分必要的。

正是上述问题的驱使,使我们萌发了探讨如何在 EXCEL 下实现多元统计分析的想法。因为,目前几乎所有的一般计算机用户中,电脑上都安装了微软的 OFFICE 办公软件,直接在电子表格 EXCEL 上进行多元统计分析,显然是十分方便的事情。而且,这些年来学术界

也已出版了一些很好的多元统计分析书籍,在这些书籍所介绍的方法原理基础上,我们想要做的就是如何使其变得更加通俗易懂和易于理解。与此同时,在过去有关计算机和多元统计分析书籍中也相继出版并发布一些基于 BASIC 计算机程序,我们想要做的就是如何在这些程序基础上,使其能够方便地在 EXCEL 上运行,这样就可以节约一大笔使用专业统计软件的费用。于是,我们把两种想法结合,并根据多元统计分析在经济研究中的应用,形成了呈现给读者您的这本《EXCEL 与多元统计分析》。

本书在完成过程中,力求做到:

第一、以解决问题为驱动,介绍多元统计分析方法原理和思想。以提出问题为出发点,以解决问题和实际应用为目的,针对财经类问题研究、财经分析应用的需要,侧重阐述多元统计分析的基本概念、基本思想、方法原理和分析步骤,而略去很多的数学证明。

第二、基于多元统计分析 BASIC 程序,以 EXCEL 工作簿进行数据维护和管理,利用 EXCEL 的 VBA 程序开发多元统计分析宏,包括:多元回归分析宏、主成分分析宏、因子分析宏、聚类分析宏、判别分析宏、对应分析宏、典型相关分析宏等,再将结果以 EXCEL 数表形式输出,最终实现在 EXCEL 中进行多元统计分析,并且把各种分析宏进行打包,最终形成应用方便的“多元统计分析程序”,随书赠送光盘,只要进行安装,在 EXCEL 的菜单上就出现了“多元统计分析程序”,应用十分方便。

第三、以经济研究为案例选择典型实际问题进行分析,以便对非统计专业财经类读者,更便于对多元统计分析方法的理解和应用。

本书主要内容包括,多元统计分析方法概述(第一章),多元正态分布(第二章),方差分析(第三章),多元回归分析(第四章);聚类分析(第五章),判别分析(第六章),主成分分析(第七章),因子分析(第八章),对应分析(第九章),典型相关分析(第十章)。本书可以作为财经类统计专业本科生和非统计专业研究生的多元统计分析教材,也可以作为广大经济理论工作者和实际部门研究人员的参考书。

本书在写作的过程中,傅德印同志负责全书的整体设计、多元统计分析算法设计和全书统稿总纂。张旭东同志负责 VBA 程序的改写、开发。其中各章完成情况为,第一章、第二章由傅德印、张旭东完成;第三章、第四章、第六章、第七章、第八章、第九章由傅德印、乔丽华、张旭东完成;第五章、第十章由傅德印、王理、张旭东完成。

本书在写作的过程中,参考了大量已有的同类书籍,以及原有多元统计分析 BASIC 程序,其中,主要的参考书目我们已经罗列在参考文献中,没有这些工作为基础,我们的书是无法完成的,在此向这些作者表示感谢。本书出版过程中也得到了中国统计出版社杨映霜女士的大力支持和帮助,也得到了一些师长和朋友们的支持和帮助,在此一并表示感谢。

由于作者水平有限,书中难免有局限性和片面性,疏漏和错误在所难免,尤其是一些自己在学习和应用过程中的体会以及设计的计算程序,可能有许多不当的地方,还恳切希望读者批评指正!

傅德印于金城
2006 年 12 月

目 录

第一章 概论	(1)
第一节 多元统计分析方法概述	(1)
一、多元统计分析研究对象	(1)
二、多元统计分析方法的主要内容	(1)
三、多元统计分析方法在经济统计中的具体应用	(2)
四、多元统计分析方法在经济统计中应用的一般步骤	(3)
第二节 多元统计分析在 Excel 中的实现	(4)
一、Excel 实现多元统计分析的必要性与可行性	(4)
二、Excel 实现多元统计分析的基本步骤和方法	(5)
三、Excel 实现多元统计分析的基本特点	(9)
四、Excel 实现多元统计分析的运行环境	(10)
五、Excel 实现多元统计分析的功能介绍	(10)
第三节 本书的内容和结构安排	(11)
一、本书主要内容安排	(11)
二、Excel 实现多元统计分析的主要操作步骤	(11)
第二章 多元正态分布	(17)
第一节 多元分布的基本概念	(17)
一、随机向量	(17)
二、多元分布函数和多元密度函数	(18)
三、多维随机向量的边缘密度、独立性与条件分布	(19)
四、多维随机向量的数字特征	(20)
第二节 多元正态分布及其参数估计	(22)
一、多元正态分布密度函数	(22)
二、多元正态分布的数字特征	(23)

三、多元正态分布的参数估计·····	(23)
第三节 多元正态分布的假设检验·····	(24)
一、三个常用的重要抽样分布·····	(24)
二、一个正态总体均值向量的检验·····	(27)
三、两个正态总体均值向量的检验·····	(27)
四、多个正态总体均值向量的检验——多元方差分析·····	(29)
五、正态总体的协方差阵检验·····	(30)
第三章 方差分析·····	(33)
第一节 方差分析的基本问题·····	(33)
一、方差分析研究的问题·····	(33)
二、方差分析的基本思想·····	(34)
第二节 单因素方差分析·····	(34)
一、单因素方差分析的计算步骤·····	(34)
二、单因素方差分析中的其他问题·····	(38)
三、单因素方差分析在 Excel 中的实现·····	(38)
第三节 双因素方差分析·····	(42)
一、双因素方差分析的类型·····	(42)
二、无交互作用的双因素方差分析·····	(42)
三、有交互作用的双因素方差分析·····	(44)
四、双因素方差分析在 Excel 中的实现·····	(46)
第四章 回归分析·····	(56)
第一节 简单回归分析法·····	(57)
一、模型和参数估计·····	(57)
二、模型的检验·····	(61)
三、进行预测·····	(65)
四、简单回归分析在 Excel 中的实现·····	(68)
五、简单回归分析在 Excel 中实现的第二种方法·····	(72)
第二节 多元线性回归分析·····	(75)
一、模型和参数估计·····	(75)

二、模型检验	(79)
三、多重共线性	(81)
四、回归模型变量子集合的选择	(83)
五、多元线性回归模型在 Excel 中的实现	(84)
第三节 逐步回归分析	(90)
一、逐步回归分析基本原理	(90)
二、引入或删除变量的依据及检验	(91)
三、逐步回归中回归系数的求解	(93)
四、逐步回归分析在 Excel 中的实现	(96)
第四节 多对多线性回归分析	(114)
一、多对多线性回归分析模型	(114)
二、多对多线性回归分析模型的参数估计	(116)
三、多对多线性回归系数向量的假设检验	(117)
四、多对多线性回归分析的计算步骤	(118)
五、多对多线性回归分析在 Excel 中的实现	(120)
第五节 岭回归分析	(125)
一、岭回归的基本思想	(125)
二、岭回归估计的性质	(126)
三、岭回归估计中回归系数的确定	(127)
四、岭回归分析在 Excel 中的实现	(127)
第六节 趋势面分析	(138)
一、趋势面分析的一般原理	(138)
二、趋势面分析中的模型参数估计	(139)
三、趋势面模型的适度性检验	(141)
四、趋势面分析的步骤	(143)
五、趋势面分析在 Excel 中的实现	(143)
第五章 聚类分析	(151)
第一节 聚类分析概述	(151)
一、聚类分析的定义	(151)
二、聚类分析的种类	(151)

三、聚类分析中样品或变量亲疏程度的测定	(152)
第二节 系统聚类分析	(158)
一、系统聚类分析的基本思想和分析步骤	(158)
二、常用系统聚类分析方法	(159)
三、系统聚类分析方法的统一公式	(167)
四、系统聚类分析方法的性质	(168)
五、Q 型聚类分析时典型变量的选择与剩余信息的剔除	(168)
六、谱系分类的确定	(169)
七、系统聚类分析在 Excel 中的实现	(169)
第三节 动态聚类法	(178)
一、动态聚类法基本思想	(178)
二、动态聚类分析的基本原理	(179)
三、动态聚类分析主要计算步骤	(181)
四、动态聚类分析在 Excel 中的实现	(182)
第四节 模糊聚类分析	(189)
一、模糊聚类分析的基本思想	(189)
二、模糊聚类分析的基本原理	(189)
三、模糊聚类分析的主要步骤	(193)
四、模糊聚类分析在 Excel 中的实现	(194)
第五节 图论聚类分析	(200)
一、图论聚类分析的基本思想	(200)
二、图论聚类分析的基本原理	(200)
三、图论聚类分析在 Excel 中的实现	(202)
第六章 判别分析	(207)
第一节 判别分析的基本原理和模型	(207)
一、判别分析概述	(207)
二、判别分析方法	(208)
三、判别分析在 Excel 中的实现	(223)
第二节 逐步判别分析	(237)
一、逐步判别分析的基本思想	(237)

二、逐步判别分析的基础理论——对判别变量附加信息的检验	(237)
三、引入和剔除变量的依据和检验统计量	(240)
四、求解判别函数中的矩阵变换	(241)
五、建立判别式,对样品判别分类	(242)
六、逐步判别分析在 Excel 中的实现	(243)
第七章 主成分分析	(253)
第一节 主成分分析的原理及模型	(253)
一、主成分分析基本思想与数学模型	(253)
二、主成分分析的几何解释	(255)
三、主成分分析的应用	(256)
第二节 主成分的导出及主成分分析步骤	(257)
一、主成分导出	(257)
二、主成分分析计算步骤	(259)
第三节 主成分分析在 Excel 中的实现	(261)
第八章 因子分析	(272)
第一节 因子分析的原理与模型	(272)
一、因子分析的基本思想和数学模型	(272)
二、因子载荷矩阵的统计意义	(274)
第二节 因子载荷矩阵的估计方法	(276)
一、因子载荷矩阵的估计	(276)
二、因子载荷矩阵的方差最大正交旋转	(278)
三、因子载荷矩阵的 Promax 协旋转	(281)
第三节 因子得分及因子分析步骤	(284)
一、因子得分	(284)
二、因子分析适用性的检验	(285)
三、因子分析计算步骤	(286)
第四节 因子分析在 Excel 中的实现	(287)

第九章 对应分析	(302)
第一节 对应分析的基本思想	(302)
一、什么是对应分析	(302)
二、对应分析的基本思想	(304)
第二节 对应分析方法的方法和原理	(304)
一、对原始数据规格变换,使 R 型和 Q 型因子分析有机结合	(304)
二、 R 型因子分析和 Q 型因子分析的对应关系	(310)
三、对应分析的计算步骤	(311)
四、对应分析在 Excel 中的实现	(312)
第十章 典型相关分析	(318)
第一节 典型相关分析的基本原理	(318)
一、典型相关分析的基本思想	(318)
二、典型相关分析的基本原理	(318)
三、典型相关分析的应用	(320)
第二节 典型变量与典型相关系数的求法	(321)
一、总体典型变量与典型相关系数	(321)
二、原始变量与典型变量之间的相关系数	(324)
三、样本典型相关变量和样本典型相关系数	(325)
第三节 典型相关系数的检验	(326)
第四节 典型相关分析计算步骤	(327)
第五节 典型相关分析在 Excel 中的实现	(328)
一、问题及背景	(328)
二、分析过程	(329)
三、输出结果	(332)
四、几点结论	(337)
参考文献	(339)
附表	(340)

第一章

概论

第一节 多元统计分析方法概述

一、多元统计分析研究对象

我们知道,统计分布是用来刻画随机变量特征及规律的重要手段,是进行统计分析的基础和前提。多元统计分析方法则是建立在多元统计分布基础上的一类处理多元统计数据方法的总称,是统计学中的具有丰富理论成果和众多应用方法的重要分支。

在现实生活中,受多个随机变量共同作用和影响的现象大量存在。统计分析中,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。一种方法是把多个随机变量分开分析,一次处理一个随机变量,分别进行研究。但是,由于变量多,无法避免地存在变量之间的相关性,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。另一种方法是同时对多个随机变量进行分析研究,即用多元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,来研究随机变量总的特征、规律以及随机变量之间的相互关系。所以,多元统计分析就是研究多个随机变量及其相互关系的统计规律性的一门学科。

二、多元统计分析方法的主要内容

近年来,随着统计理论研究的不断深入,多元统计分析方法的内容不断丰富。其中,主体内容包括多元正态总体参数估计、假设检验和常用的多元统计方法。多元正态总体参数估计、假设检验是多元统计推断的基础和核心,而常用多元统计分析方法则是具体应用。从形式上,常用多元统计分析方法可划分为两类:

一类属于单变量常用的统计方法在多元随机变量情况下的推广和应用,如多元回归分析,典型相关分析等;

另一类是对多元变量本身进行研究所形成的一些特殊方法。如主成分分析,因子分析,聚类分析,判别分析,对应分析等。

具体来说,常用的多元统计分析方法主要包括:多元数据图表示法、多元回归分析、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、多维标度法等。

三、多元统计分析方法在经济统计中的具体应用

多元统计分析方法的应用十分广泛。多元统计分析方法可以应用于地质科学、气象科学、医疗卫生、体育、语言学、考古学、教育学、心理学以及经济学、管理学等各个学科领域。其中,仅就在经济学和管理学中的应用,主要可集中在如下的场合或领域:

1. 对多个变量需要进行降维处理,选择数目较少的变量子集合。如在经济统计分析中,当对地区、产业或企业经济效益进行评价分析时,对地区经济发展水平评价,对地区或企业竞争力进行对比,对人民生活质量等进行综合评价分析时,为了能够全面刻画所要研究对象的数量特征,往往要调查、搜集该研究对象的多方面的统计数据,这样就构成了多元的统计数据,统计数据的维数越多,所反映的内容就越丰富、越全面,包括的信息量越大,但另一方面,统计数据的维数越多,数据就越复杂,而且由于各个变量之间的相关性,也给统计分析带来困难,因此,为了使统计分析简化,就需要在不损失太多的原有信息量的情况下,对原始数据进行降维处理,并选择相关性较小的或者互不相关的代表性变量进行分析,进而达到简化数据结构的同时,又能够进行客观分析的目的。能够进行这样数据处理的多元统计分析方法主要有,主成分分析、因子分析、对应分析等。当然主成分分析、因子分析、对应分析等方法,还具有更多的其它应用,具体到各章时再介绍。

2. 对研究对象需要进行分类研究、分类处理、构造分类模式。我们知道,统计分组可以深化人们的认识,因此进行经济统计分析时,分组研究是常用的方法之一。但是统计分组有时比较简单,比如人口总体按照性别可以分为男性人口和女性人口两组,有时统计分组是非常困难的事情,如根据各地区的经济发展水平、经济发展特征对我国各地区的经济发展类型进行划分,就不可能只根据行政区划或地理位置来简单的划分。这时就需要根据反映各地区经济发展水平、经济发展特征的多项指标所描述的内容,测算各地区的经济发展的相似程度,根据

相似程度对各地区的经济发展类型进行划分。这一过程就需要对多元数据进行分类研究、分类处理,甚至有时需要构造分类模式,用来处理这一类问题的多元统计分析方法主要是聚类分析、判别分析等。

3. 建立经济模型和利用模型进行外推。经济模型一般是指把经济变量之间的依存关系通过数学表达式的形式加以模拟。在经济模型的基础上,一方面是根据这种数学表达式所呈现出来的经济变量之间的内部结构,进行各种分析如因素分析、边际分析或弹性分析;另一方面是根据自变量的变化来预测因变量的变化。在多元统计分析应用中有两大类模型,一类是预测模型,另一类是描述性模型。对于预测模型通常采用回归分析方法解决,如果经济预测变量是一个因变量,可应用多元线性回归、逐步回归、非线性回归分析方法处理;如果经济预测变量是多个因变量,可以应用多个因变量对多个自变量的多对多回归分析处理。对于描述性模型,通常采用聚类分析等方法处理。

4. 研究经济现象之间相互关系。经济现象之间的联系是普遍存在的。我们知道,简单直线相关系数可以用来刻画两个随机变量之间线性相关的程度及方向。但是,当我们要研究两组变量之间的相关程度时,如研究一组反映投入变量与另外一组反映产出变量之间的相关程度时,这时只用简单直线相关系数是不够的,在多元统计分析中,用典型相关分析可以处理两组变量之间的相关程度的分析和测量。

这里需要进一步说明的是,由于社会经济现象的复杂性以及每一种多元分析方法特殊的应用场合和自身的局限性,所以在进行经济统计分析时,有必要把多种多元统计分析方法结合运用,如主成分分析与回归分析的结合运用,因子分析与回归分析的结合运用,因子分析与聚类分析的结合运用等,这样会收到更好的效果。

四、多元统计分析方法在经济统计中应用的一般步骤

与一般统计分析方法一样,进行多元统计分析也要经过建立模型、进行参数估计、假设检验以及预测控制等步骤,具体是:

1. 根据经济理论进行定性分析,设计理论模型;
2. 对实际经济活动的现象抽取样本,并取得样本统计资料;
3. 对描述样本的指标利用统计分析方法(如相关分析、主成分分析等)进行统计分析,选择最佳的统计指标;
4. 根据最佳指标的样本数据,估计参数,建立数量模型;
5. 对数量模型进行检验、优化以及运用。

第二节 多元统计分析在 Excel 中的实现

一、Excel 实现多元统计分析的必要性与可行性

由于多元统计分析方法要分析大量的多维数据,进行数据处理时在计算上是非常复杂的,为此,应用多元统计分析方法都离不开计算机软件的使用。目前,可以用于多元统计分析的软件很多,如常用的有 SPSS, SAS, Statistica, Matlab, S-plus, NCSS 等,这些软件都是国际上著名的统计软件,而且能够很好的实现常用的多元统计分析方法,所以,对推动多元统计方法在实际中的广泛应用起到了重要作用。

但是,这些软件都属于专业性很强的统计软件,就我国目前使用情况看,还主要集中于专业统计人员的使用,对于非统计专业的人员来说,有时由于手头不具备专业统计软件,往往又会限制了多元统计分析方法的应用,推广面很窄。另外,专业统计软件的使用费用都较高,购置一套软件少则几千人民币,多则上万,而且有的还有应用期限限制。这些都不利于多元统计分析方法广泛应用。

Excel 是微软公司出品的 Office 系列办公软件中的一个组件,确切地说,它是一个功能强大、技术先进、使用方便的电子数据表格软件。它可以进行各种数据处理、统计分析和辅助决策操作,广泛应用于管理、统计、财政、金融等众多领域。而且不断更新的版本,又不断地加入新的功能,使其数据分析的能力不断改进,提高了工作的效率。更为重要的是,在我国,广泛使用 Excel 办公软件已经成为不争现实,可以说,目前我国办公用的计算机上绝大多数都安装了 Excel 软件。Excel 软件的使用普及率要远远大于专业统计软件的使用率,为此,加强 Excel 软件中统计分析功能的开发和使用,对推动统计方法的广泛应用具有重要作用且富有成效。

诚然,如果我们开发基于 Excel 软件的使用方便并且费用低廉的多元统计分析软件,将不仅会推动常用统计方法的应用,更有助于推动多元统计分析方法的普及应用。

目前,Excel 软件中已经具备可以进行统计分析的一组数据分析工具,称为“分析工具库”。利用“分析工具库”,即可进行常用的统计分析。只需为每一个分析工具提供必要的数据和参数,该工具就会使用适宜的统计或工程函数,在输出表格中显示相应的结果。其中有些工具在生成输出表格时还能同时生成图表。具体“分析工具库”,所能进行的常用统计分析包括:方差分析、相关系数分析工具、协方差分析、描述统计、指数平滑、傅立叶分析、F - 检验、双样本方差、