

T 高等院校研究生用书  
EXTBOOKS FOR  
GRADUATES

# 现代统计分析 方法与应用 (第四版)

何晓群 编著

 中国人民大学出版社

T 高等院校研究生用书  
EXTBOOKS FOR  
GRADUATES



**现代统计分析  
方法与应用** (第四版)

何晓群 编著

中国人民大学出版社

· 北京 ·

图书在版编目 (CIP) 数据

现代统计分析方法与应用/何晓群编著. —4 版. —北京: 中国人民大学出版社, 2016. 11  
高等院校研究生用书  
ISBN 978-7-300-23662-9

I. ①现… II. ①何… III. ①统计分析-研究生-教材 IV. ①C812

中国版本图书馆 CIP 数据核字 (2016) 第 279665 号

高等院校研究生用书  
现代统计分析方法与应用 (第四版)  
何晓群 编著  
Xiandai Tongji Fenxi Fangfa yu Yingyong

---

出版发行	中国人民大学出版社		
社 址	北京中关村大街 31 号	邮政编码	100080
电 话	010-62511242 (总编室)		010-62511770 (质管部)
	010-82501766 (邮购部)		010-62514148 (门市部)
	010-62515195 (发行公司)		010-62515275 (盗版举报)
网 址	<a href="http://www.crup.com.cn">http://www.crup.com.cn</a> <a href="http://www.ttrnet.com">http://www.ttrnet.com</a> (人大教研网)		
经 销	新华书店		
印 刷	北京市鑫霸印务有限公司	版 次	1998 年 5 月第 1 版 2016 年 11 月第 4 版
规 格	185 mm×260 mm 16 开本	印 次	2016 年 11 月第 1 次印刷
印 张	19.25 插页 1	定 价	39.00 元
字 数	450 000		

---

版权所有 侵权必究 印装差错 负责调换



## 第四版前言

自 1998 年本书出版以来,承蒙数万读者的厚爱,许多高校都采用本书作为人文社科、经济管理类研究生教材。有许多教师和学生给予我热情的鼓励和中肯的批评,这都是读者对我及本书的无私关心和奉献。在此我谨表衷心感谢。

几年来,一种责任心总在督促我对本书作出修改,重新出版,以表达对读者的谢意。在中国人民大学出版社王伟娟女士的精心策划下,我在第三版的基础上进行了一些修改。

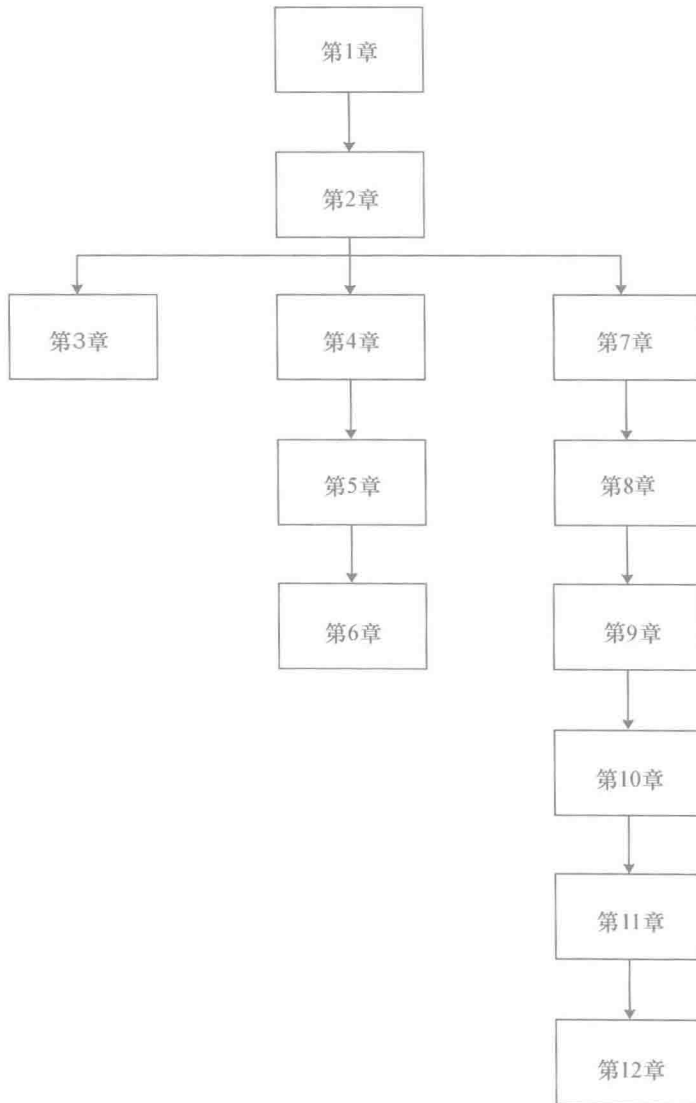
本次再版,我对本书做了如下修改:

1. 采用最新数据更换了一些例题。
2. 大部分例题都采用 SPSS (社会科学专用统计软件) 或 SAS 进行分析。
3. 对第三版书中一些错误之处进行了修正,并且斟酌了书中的一些文字。
4. 由于多变量作图在许多基础统计学教材和软件中都已描述,故本版删掉了多变量图示法。
5. 为方便读者,将本书所有例题和习题的数据都做成了电子文件,读者可在 [www.ruc-6sigma.com](http://www.ruc-6sigma.com) 网站或人大经管图书在线 ([www.rdjg.com.cn](http://www.rdjg.com.cn)) 的《现代统计分析方法与应用》数据文件夹下载。

本书总的参考教学课时为 54 学时。根据笔者的经验,学生掌握这些基本方法和技能并不困难。如有计算机配合,教学效果将会更好。选用本书的教师在讲授时可有一定的灵活性,根据专业的不同有选择地讲授本书内容。一般 2 学分的课程可只讲授前 6 章的内容。从第 7 章开始是多元统计分析的内容,可看成是高级统计分析方法。下面是本书建议讲授或阅读顺序图。

路线 1: 对欲掌握基本统计分析方法者,27 学时可选讲第 1 章、第 2 章、第 4 章、第 5 章、第 6 章。

路线 2: 对欲掌握较为高级统计分析方法者,27 学时可选讲第 7 章、第 8 章、第 9



章、第 10 章、第 11 章、第 12 章。

第 3 章可根据需要任意选择在两条路线中讲授。

中国人民大学六西格玛质量管理研究中心

何晓群

2016 年 10 月



## 第一版前言

随着社会主义市场经济体制在中国的快速推进,许多非常复杂和重要的社会、经济问题急切地需要我们去研究,如金融风暴、通货膨胀、税务稽查、保险精算、环境保护、可持续发展、国际竞争、公众意向,传统的定性研究理论和方法已无法满足这些热点问题的研究需要。21世纪,深刻的社会变革、迅猛的经济发展,使我国的人文社会科学、财经管理类研究生面临严峻的挑战和难得的机遇。时代呼唤我们精通定量分析的研究方法,掌握定性定量有机结合的研究技能。《现代统计分析方法与应用》一书正是适应这一需要,为非统计专业的人文社会科学、财经管理类硕士研究生学习现代统计分析方法而编著的。

统计方法与技术是现代社会的经济研究运用的基本方法。自1969年设立诺贝尔经济学奖以来,已有40多位学者获奖。这些获奖者大都精通现代统计方法,对统计方法的运用极为娴熟,在社会经济研究中取得了举世瞩目的成就。学习和运用统计方法已成为时代对我们的要求。

作者假定学生已具有线性代数、概率论与数理统计的基础知识,本着提高研究生量化分析能力的宗旨,在众多统计方法中,仅选择部分最实用的现代统计分析方法。在不失理论严密性的前提下,力求将问题的背景、方法的思想、具体的步骤、分析的技巧讲清楚。为重点突出方法的应用,每种方法都尽可能结合中国社会、经济、管理方面的实际问题,以案例研究为导向,为非统计专业研究生进行量化分析起一定示范作用。

1996年,中国人民大学率先在非统计专业的人文社会科学、财经管理类研究生中开设“统计方法与技术”必修课,作者有幸给中国人民大学96级、97级硕士研究生和MBA主讲此课。在教学实践中,学生们给了我许多启发和鼓励,因为他们结合自己的专业,对统计方法的学习产生了浓厚的兴趣,看到了统计方法的用武之地,清楚哪些方法最有用;他们在学习的过程中也渴望拥有一本合适的教材。

本书的大部分内容都给非统计专业研究生讲授过,根据笔者的经验,如有计算机配合,学生掌握这些基本方法和技能并不困难。选用本书的教师可有一定的灵活性,根据不同专业有选择地讲授本书内容。本书参考教学课时为54学时。



本书也可作为经济统计专业现代统计分析方法课的教材。此书还可作为从事社会、经济、管理等研究和实际工作的同志进行量化研究的参考书。

本书在写作过程中,始终得到中国人民大学研究生教材工作委员会和林岗、袁卫两位副校长的关注。中国人民大学研究生院高培勇教授、统计系赵彦云教授,经济科学出版社高续增同志给予作者许多鼓励和支持。香港浸会大学数学系首席教授方开泰先生对本书的写作给予悉心指点。中国人民大学统计系兼职教授、中国科学院院士陈希孺教授,上海财经大学张尧庭教授对本书的具体内容提出过不少中肯建议。中国人民大学统计系博士孟生旺、刘勤、宋光辉、刘长标等为本书的部分案例做过一些验算。西安统计学院王国梁教授也曾给作者许多热情鼓励。在此,我谨向对本书出版给予支持的师长和朋友表示衷心的感谢。

由于本人学识有限,书中谬误之处在所难免,恳请读者批评指正。

何晓群

1998年5月1日于中国人民大学静园



# • 目 录 •

第 1 章 概论 .....	1
§ 1.1 为统计学正名 .....	1
§ 1.2 市场呼唤真统计 .....	2
§ 1.3 统计分析方法及其应用概述 .....	4
思考与练习 .....	10
第 2 章 统计学基础回顾 .....	11
§ 2.1 统计数据的整理与描述 .....	11
§ 2.2 几种重要的概率分布 .....	14
§ 2.3 多元分布的基本概念 .....	19
§ 2.4 多元正态分布 .....	23
§ 2.5 参数估计 .....	28
§ 2.6 假设检验 .....	30
思考与练习 .....	34
第 3 章 定性数据的 $\chi^2$ 检验 .....	35
§ 3.1 多项分布与 $\chi^2$ 检验 .....	35
§ 3.2 列联表分析 .....	39
§ 3.3 一致性检验 .....	47
§ 3.4 拟合优度检验 .....	49
思考与练习 .....	53
第 4 章 一元线性回归 .....	55
§ 4.1 一元线性回归模型 .....	55





§ 4.2	回归参数 $\beta_0, \beta_1$ 的估计	59
§ 4.3	最小二乘估计的性质	62
§ 4.4	回归方程的显著性检验	64
§ 4.5	残差分析	72
§ 4.6	预测和控制	74
§ 4.7	建模总结和应注意的问题	78
	思考与练习	83
<b>第 5 章</b>	<b>多元回归模型</b>	<b>86</b>
§ 5.1	多元线性回归模型	86
§ 5.2	回归参数的估计与检验	89
§ 5.3	自变量选择与逐步回归	100
	思考与练习	106
<b>第 6 章</b>	<b>定性数据的建模分析</b>	<b>109</b>
§ 6.1	对数线性模型基本理论和方法	109
§ 6.2	对数线性模型分析的上机实现	111
§ 6.3	Logistic 回归基本理论和方法	115
§ 6.4	Logistic 回归的建模总结	123
	思考与练习	124
<b>第 7 章</b>	<b>聚类分析</b>	<b>125</b>
§ 7.1	聚类分析的基本思想	125
§ 7.2	相似性度量	128
§ 7.3	类和类的特征	131
§ 7.4	聚类方法	133
§ 7.5	实际例子	141
	思考与练习	149
<b>第 8 章</b>	<b>判别分析</b>	<b>151</b>
§ 8.1	判别分析的基本思想	151
§ 8.2	距离判别	152
§ 8.3	贝叶斯判别	154
§ 8.4	费歇判别	155
§ 8.5	逐步判别	156
§ 8.6	判别分析应用的几个例子	157
	思考与练习	171



第 9 章 主成分分析 .....	173
§ 9.1 主成分分析的基本思想 .....	173
§ 9.2 总体主成分及其性质 .....	176
§ 9.3 由样本数据求主成分 .....	183
§ 9.4 主成分分析步骤及框图 .....	184
§ 9.5 主成分分析的应用 .....	185
思考与练习 .....	200
第 10 章 因子分析 .....	201
§ 10.1 因子分析的基本思想 .....	201
§ 10.2 因子载荷的求解 .....	205
§ 10.3 因子分析的上机实现 .....	210
思考与练习 .....	231
第 11 章 对应分析 .....	232
§ 11.1 对应分析的基本理论 .....	232
§ 11.2 对应分析的步骤及逻辑框图 .....	238
§ 11.3 对应分析的上机实现 .....	239
思考与练习 .....	253
第 12 章 典型相关分析 .....	254
§ 12.1 典型相关分析的基本理论 .....	254
§ 12.2 典型相关分析的上机实现 .....	260
思考与练习 .....	277
参考文献 .....	278
附录 .....	280
表 1 泊松分布表 .....	280
表 2 标准正态分布表 .....	282
表 3 正态分布分位数表 .....	285
表 4 卡方 ( $\chi^2$ ) 分布表 .....	288
表 5 相关系数检验表 .....	289
表 6 $t$ 分布的临界点 .....	290
表 7 $F$ 分布表 .....	291



## • 第 1 章 •

# 概 论

为使读者对现代统计学及其应用有较为系统的认识，本章将对统计学及其学科性质，统计学在现代社会、经济和管理等方面的应用给予扼要论述，并对统计分析方法与技术应用的步骤流程进行描述。

### § 1.1 为统计学正名

人文社会科学、财经管理类研究生在本科阶段大都学过“社会经济统计学原理”或“统计学”课程。“统计”一词对中国的普通老百姓来说也并不陌生。不少统计学者更是认为政府统计才是统计。1988年出版的《中国大百科全书》经济卷Ⅱ，把统计学定义为一门社会科学。《大英百科全书》认为统计学是一门收集数据、分析数据，并根据数据进行推断的艺术和科学。然而，究竟什么是统计学，统计学的学科性质是什么，我们将要学习的统计学理论和方法究竟是社会科学，还是自然科学，同样需要我们去认识和思考。

1990年以前，中国的高等统计教育，尤其是财经类的统计专业，基本上是苏联1954年统计工作会议决议的那种模式，即把统计学定义为一门具有阶级性、党性的社会科学。这种统计理论体系被中国统计界的一些学者称为“马克思主义的无产阶级统计学”。令人遗憾的是至今有不少统计界学者仍持这种观点。统计学是否具有阶级性我们姑且不论，但苏联统计模式给中国统计界提供的示范作用以及对统计学在中国的发展所产生的历史性不良影响是不容置疑的。

在中国，人们把统计学确立为社会科学及其经济学的的一个子学科后，不仅在科学概念上把社会经济统计学和数理统计学割裂开来，更把数理统计当作资产阶级货色加以批判（并非只有“文革”时期），不幸的是社会经济统计学把自己也限定在以简单描述为工具的政府统计圈子里。近半个世纪在中国作为显学的社会经济统计学对中国高度集中的政府计划经济起了一定作用，但无法否认的是这种学科模式也拉大了中国统计与世界统计水平之间的差距。

描述统计是对所收集的数据资料进行加工整理、综合概括，通过图示、列表和数对资



料进行分析和描述。推断统计是在收集、整理观测样本数据的基础上，对有关总体作出推断，其特点是根据随机性的观测样本数据以及问题的条件和假定，对未知事物作出的以概率形式表述的推断。

一门学科从定性的描述进入定量的分析和计量，标志着这门学科达到了比较成熟的阶段。从描述统计学发展到推断统计学已成为统计学发展成熟的重要标志。不难想象，统计学如果只有描述理论和方法，肯定是发育不全的统计学。也正是由于推断统计的产生，才使人类的科学技术借助统计学理论取得了一个又一个重大成就。著名的基因论之所以不同于以往对遗传现象的描述，就在于它是根据两两具有不同性状的个体杂交实验所获得的大量数据，运用数理统计方法推导出来的，从而为进一步发展遗传理论奠定了基础。

改革开放对中国的统计学来说是挑战更是机遇。严峻的挑战和深刻的矛盾要求统计界必须坚持实事求是的科学态度。经过一大批统计学家的艰苦努力，我国统计界终于发生了质的变化和飞跃。1992年11月，国家技术监督局正式批准统计学为一级学科，国家标准局颁布的学科分类标准已将统计学列为一级学科，1998年教育部进行的本科专业调整也将统计学归为理学类一级学科。一级学科的地位表明统计学既不是数学的子学科，也不是经济学的子学科，统计学就是统计学。统计学一级学科的地位表明中国统计在与国际接轨的进程中迈出了重要一步。在国际惯例中根本就不存在社会经济统计学和数理统计学之分。统计学（statistics）是一门综合性很强的边缘学科，它不是数学，更不是经济学。统计学是研究客观事物数量关系和数量特征的方法论学科。既然是处理数据的方法论学科，因而哪个阶级都可以利用它，它是人类共同的财富。

统计学到底应该隶属于哪一个学科，中国统计学界今天还对这个不应争论的问题激烈地争论着，是理学，经济学，还是管理学？无论怎样，笔者认为总比放在经济学里要合理得多。

需要注意的是，本书所介绍的统计学与过去的社会经济统计学相比有着很大的不同。这个差别不仅体现在内容的深度和广度上，更体现在处理问题的思想观点和方法上。在统计学的学习和应用中将充分体现从具体到一般的概念，再从一般概念回到具体事物上去的这种辩证观点和较为严格的逻辑推理。统计学给实际问题的应用带来新思维，应用给统计学增添活力。统计学的生命力就在于应用。对此问题有兴趣者可进一步阅读书末的参考文献 [1]。

## § 1.2 市场呼唤真统计

统计学在中国被列为一级学科，这只是中国统计与国际接轨的第一步。由于统计界长期受苏联模式的禁锢，与中国其他人文社会科学一样，统计学面对市场经济已显露出日趋严重的危机。

中国目前仍处在新旧经济体制的变革时期，面对市场经济，任何学科都需要调整自己的位置，重新审视和制定自己的发展目标。市场经济的理论和实践对统计信息的需求急剧增加，对统计学理论和方法提出了更高的要求，苏联模式的统计学面对市场经济的需求已



显得无能为力。现代市场经济的理论与时代的潮流使中国人文社会科学、财经管理类学者的弱点暴露无遗。中国几千年的优秀文化会给中国学者的思辨性思维以深厚的底蕴，但也留下了忽视形式逻辑的弊端。先秦诸子以来，我国学者的思辨性思维是较为发达的，但这些思辨性思维常常缺乏经验事实做基础，空泛的议论较多，实质性的内容较少，逻辑推理的思维方式淡化，更不追求严密的公理化体系，以至于我国人文社会科学、财经管理类学者面对现代社会经济的发展有力不从心之感。

在 21 世纪的进程中，如何摆脱困境、走向未来，置身于古老东方文化氛围之中的中国学者需要冷静反思。时代要求我们必须抛开偏见，克服由于文化背景的差异所产生的观念屏障，正确理解与批判地吸收建立在发达商品经济基础上的外来文化，加强数学方法、统计学方法的学习，提高我们定性分析与定量分析相结合的能力。这样，在新世纪的实践进程中，中国人才会大步赶上世界发达国家。现在我们在人文社会科学、财经管理类研究生中开设统计方法与技术的课程，就是向前迈进的一种有力举措。

哪里有数据，哪里就有统计。统计学是总结经验的学科。统计方法可以帮助我们获得利益，统计方法可以保护我们已获得的利益。统计数据中隐含着非常丰富的宝贵信息，如果用简单的描述统计方法，就只是对统计数据的粗加工，难以看到数据中隐含的规律性。要从杂乱无章的统计数据中发现事物物质的规律性就像雾里看花，只有靠现代统计分析方法才能真切地看到其统计规律性。

人文社会科学、财经管理类的学者要跟上时代的步伐，就必须在研究事物质的规律性的同时，研究其量的规律性。任何现象都是质和量的统一，质是量的基础，而一定的量是质存在的必要条件。定性分析是定量分析的基础，定量分析则使定性分析更加准确可靠，使人们对质的规律性的认识更加深入全面。在马克思主义理论的指导下，在人文社会科学、财经管理类研究中融入量化分析，是学科本身进一步完善的表现。

当今中国不仅需要经济理论家，更需要高级经济管理人才。在世界经济贸易中，随着 ISO 9000 系列标准的形成和发展，质量体系认证在世界各国兴起，20 世纪末一股 ISO 9000 热席卷全球，质量体系认证已经成为国际贸易中衡量供方质量保证能力和水平的标志。对统计技术在质量管理中的作用，最早提出全面质量管理理论的美国通用电气公司的质量管理部部长费根堡姆博士曾说：“在全面质量管理活动中，无论何时、何地都会用到统计方法，但是统计方法只是全面质量管理的一个内容，它不等于全面质量管理。”“统计方法所表达的观点，对全面质量管理的整个领域都有深刻影响。”ISO 9000 系列标准的基本点是要求企业在生产过程中建立有效的质量保证体系，并对质量体系中相互关联、相互作用的若干要素进行有效的控制。在过程质量控制中，科学、有效的方法之一就是统计方法。因此，在 ISO 9000 系列标准的各个模式中以及质量管理和质量体系要素指南中，都要求在市场分析、产品设计、工序控制、性能评定、数据分析等方面广泛使用统计技术，其范围包括实验设计、方差分析、显著性检验、累计和控制图、抽样检验等技术。在今天的管理界被普遍认可的六西格玛管理更是统计学思想和方法的典型应用。

现代经济学一个很重要的标志就是模型技术的应用，而这里的模型技术更多的是指统计模型技术。诺贝尔经济学奖获得者萨缪尔森曾说，第二次世界大战以后的经济学是计量经济学的时代。计量经济学的模型技术就是以统计模型为基础的。自 1969 年设立诺贝尔

经济学奖以来,已有近 80 位学者获奖,其中 2/3 以上获奖者是统计学家、计量经济学家、数学家。从大多数获奖者的经历和著作看,他们对统计方法的运用具有娴熟的技巧。瑞典皇家科学院爱立克·伦德伯教授在诺贝尔经济学奖的授奖仪式上精辟地指出:“你们都是把经济学发展为数学和定量科学的先行者,你们借助于发展成熟的理论和统计分析来创造经济政策和经济规划的合理基础的贡献,涉及重大科学突破。”这足以说明统计量化分析研究已成为现代经济科学研究的重要手段。

在 21 世纪的进程中,中国的人文社会科学、财经管理类学者肩负着时代的重托。社会和谐发展,社会主义新农村建设,金融、投资风险,人口与社会保障,经济持续增长,环境保护问题等,都迫切地等待着我们去深入研究。这些问题的研究都将借助于统计方法与技术的应用而获得成功。

### § 1.3 统计分析方法及其应用概述

从高斯(C. F. Gauss)描述天文观测的误差而引进正态分布,并使用最小二乘法作为一种估计方法至今,现代统计的发展已有 200 多年的历史。统计学在分析数据、探索数据规律性、研究现实问题中已形成许多各具特点的思想方法。除本科阶段我们学过的一些基本统计方法与技术外,本书主要介绍一些最实用的统计分析方法。

现代统计分析方法从研究问题的角度大致可划分为四大类:分类分析方法、结构简化方法、相关分析方法、预测决策方法。我们可将上述四类及其具体方法用树形图表示(见图 1-1)。



图 1-1

下面我们进一步对统计分析方法应用的步骤及其流程进行描述(见图 1-2)。

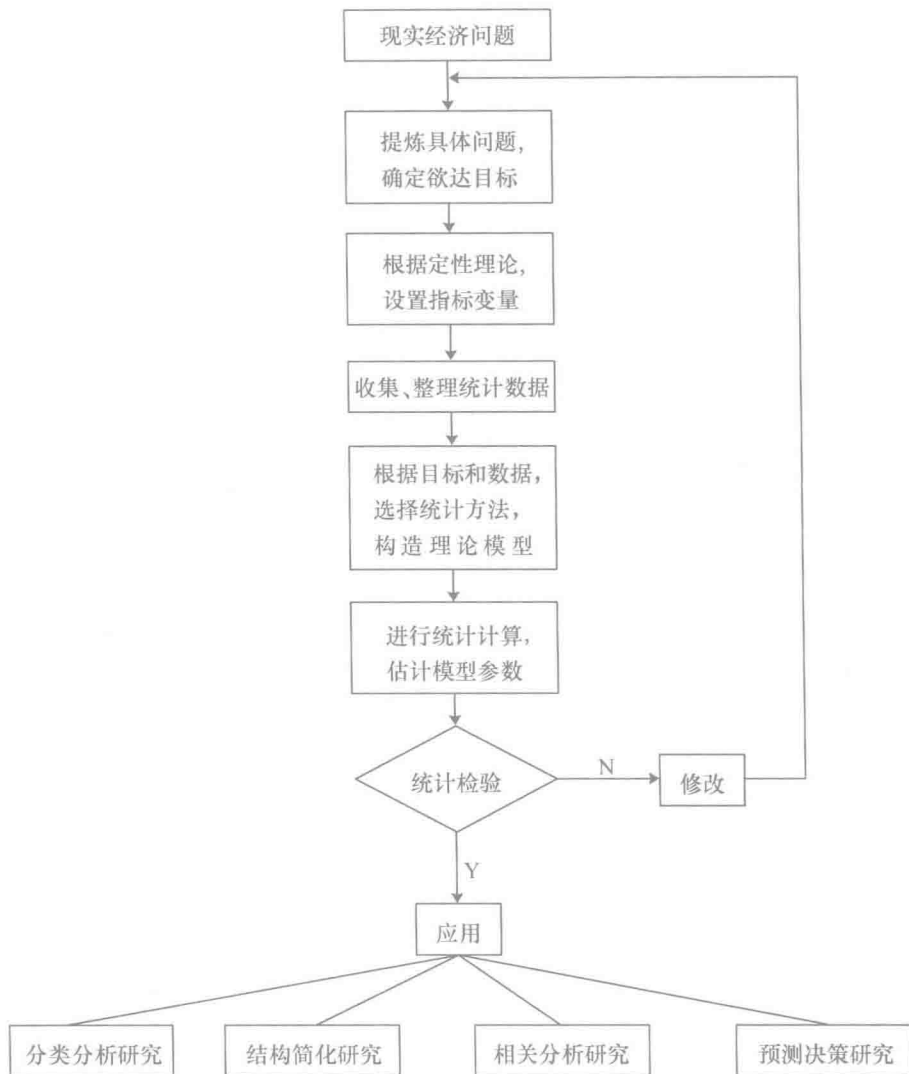


图 1—2 统计分析方法应用步骤流程图

### 一、提炼具体问题，确定欲达目标

人文社会科学、财经管理等领域存在大量需要深入探讨的实际问题，它们往往表现为一个立体网络体系，变量之间的关系复杂多样。然而能力、知识和条件的局限使得我们不可能对它们的研究面面俱到，因此，在实际问题的研究中，集中精力抓主要矛盾、提炼问题、确定切实可行的目标，就成为首要步骤。

### 二、根据定性理论，设置指标变量

面对具体的实际问题，根据定性理论进行分析研究，设置指标变量是统计方法研究实际问题的重要一环。如回归分析方法主要是揭示事物间变量联系的因果关系，在研究某一



经济问题时，我们必须根据具体经济问题的研究目的，利用经济学理论，从定性角度来确定这种经济问题中各因素之间的因果关系。当我们把某一经济变量作为“果”之后，更重要的就是正确选择作为“因”的变量。在经济问题回归模型中，前者称为“内生变量”或“被解释变量”，后者称为“外生变量”或“解释变量”。变量的选择，关键在于能否正确把握所研究经济活动的经济学内涵，这就要求研究者对所研究的经济问题及其背景应有足够了解。例如，要研究中国通货膨胀问题，必须懂得一些金融理论。通常把全国零售物价总指数作为衡量通货膨胀的重要指标，那么，全国零售物价总指数作为被解释变量，影响全国零售物价指数的有关因素就作为解释变量。书末的参考文献 [2] 在研究中国通货膨胀问题时，曾把国民收入、居民存款、工农业总产值、国有单位固定资产投资、货币流量、职工平均工资、社会商品零售总额等 18 个指标确定为解释变量。书末的参考文献 [3] 在研究中国储蓄波动机理时，曾把各项银行存款作为被解释变量，把货币发行量、全国零售物价指数、股票价格指数、银行利率、国债利率、居民收入等 16 个指标确定为解释变量。

对一个具体的经济问题，当研究目的确定之后，被解释变量很容易确定。被解释变量一般直接表达、刻画研究的目的，就像书末的参考文献 [2] 和 [3] 中根据研究的问题，直接将全国零售物价总指数和银行存款分别作为通货膨胀问题和储蓄问题的被解释变量。而对被解释变量有影响的解释变量的确定就不太容易。一是由于我们认识上的局限性，可能并不知道对被解释变量有重要影响的因素。二是为了模型参数估计的有效性，设置的解释变量之间应该是不相关的，而我们很难确定哪些变量是相关的，哪些是不相关的，因为在经济问题中很难找到影响同一结果的一些因素之间是独立的，这就看我们如何在多个变量中确定几个重要的且不相关的变量。三是应该引进我们从经济关系角度考虑非常重要的变量，但是在实际中并没有这样的统计数据。这一点，在建立经济模型时会经常遇到。这时，可以考虑用相近的变量代替，或者由其他几个指标复合成一个新的指标。

在选择变量时要注意与一些专门领域的专家合作。研究金融模型，要与金融专家和具体业务人员合作；研究粮食生产问题，要与农业部门的专家合作。这样做可以帮助我们确定模型变量。

另外，不要认为一个经济模型所涉及的解释变量越多越好。一个经济模型，如果把一些主要变量漏掉肯定会影响模型的应用效果，但如果鸡毛蒜皮一起进入模型也未必就好。引入的变量太多，就有可能选择一些与问题无关的变量，还可能由于一些变量的相关性很强，所反映的信息有较严重的重叠，就会出现共线性问题。变量太多时，计算工作量过大，计算误差累积也大，估计出的模型参数精度自然不高。

总之，指标变量的确定是一个非常重要的问题，是运用统计模型最基本的工作。这个工作一般一次并不能完全确定，通常要经过反复试算，最终找出最适合的一些变量。在计算机的帮助下，这已不再困难了。

### 三、收集、整理统计数据

统计模型的建立基于指标变量的样本统计数据。确定好指标变量之后，就要对这些变量收集、整理统计数据。数据的收集是建立经济问题统计模型的重要一环，是一项基础性





工作, 样本数据的质量如何对统计模型的应用水平有至关重要的影响。

常用的样本数据分为时间序列数据和横截面数据。

顾名思义, 时间序列数据就是按时间顺序排列的统计数据。如新中国成立以来历年的工农业总产值、国民收入、发电量、钢产量、粮食产量等, 每年都有一个对应的数据, 那么到 2017 年, 每种指标都有 68 个按时间顺序排列的数据, 这些都是时间序列数据。研究宏观经济问题, 有关的时间序列数据来自国家统计局或一些专业部委的统计年鉴。如果研究微观经济现象, 如研究某企业的产值与能耗, 那么数据就要在这个企业的计划统计部门获取。

对于收集到的时间序列资料要特别注意数据的可比性问题和数据的统计口径问题。如历年的国民收入数据, 是不是按可比价格计算的。中国在改革开放前, 几十年物价不变, 而从 20 世纪 80 年代初开始, 物价变动很频繁, 那么你所获得的数据是否具有可比性就需要认真考虑。在宏观经济研究中, 国内生产总值 (GDP) 与国民生产总值 (GNP) 二者在内容上是一致的, 但在计算口径上不同。国民生产总值按国民原则计算, 反映一国常住居民当期国内外所从事的生产活动; 国内生产总值则以国土为计算原则, 反映一国国土范围内所发生的生产活动。对于没有可比性和统计口径计算不一致的统计数据要进行认真调整, 这个调整过程就是数据整理的一种过程。

时间序列数据容易产生模型中随机误差项的序列相关, 这是因为许多经济变量的前后期之间总是关联的。如在建立需求模型时, 人们的消费习惯、商品短缺程度等具有一定的延续性, 它们对相当一段时间的需求量有影响, 这样就产生随机误差项的序列相关。对于具有随机误差项序列相关的情况, 要通过对数据的某种计算整理来消除序列相关性, 最常用的处理方法是差分方法。我们将在后面的章节中详细介绍。

横截面数据就是在同一时间截面上的统计数据。如同一年在不同的地块上做的施肥量与小麦产量实验的统计数据就是截面数据。再如某一年的全国人口普查数据、工业普查数据, 同一年份全国 35 个大中城市物价指数等都是截面数据。用截面数据作样本时, 容易产生异方差性。这是因为一个统计模型往往涉及众多解释变量, 如果其中某一因素或一些因素随着解释变量观测值的变化而对被解释变量产生不同影响, 就产生异方差性。如在研究城镇居民收入与购买消费品的关系时, 用  $x_i$  表示第  $i$  户的收入量,  $y_i$  表示第  $i$  户的购买量, 购买量回归模型为:

$$y_i = \beta_0 + \beta_1 x_i + u_i$$

在此模型中随机项  $u_i$  就具有不同的方差。因为在购买行为中, 低收入家庭购买差异比较小, 大都购买生活必需品; 高收入家庭购买行为差异就很大, 高档消费品很多, 选择余地很大, 这样二者之间购买物品所花费用差异就较大。因而, 用随机获取的样本数据建立回归模型, 它的随机项  $u_i$  就具有异方差。

对于具有异方差性的建模问题, 数据整理就要注意消除异方差性, 这常与模型参数估计方法结合起来考虑。我们将在后面的章节中详细介绍。

不论是时间序列数据还是横截面数据的收集, 样本容量的多少一般要与设置的解释变量数目相配套。通常为了使模型的参数估计更有效, 要求样本容量  $n$  大于解释变量个数