

中国人民大学硕士研究生系列教材

# 现代统计分析 方法与应用

● 何晓群 编著

中国人民大学出版社

# 现代统计分析 方法与应用

● 杨晓丹 编著

清华大学出版社

中国人民大学  
硕士研究生系列教材

# 现代统计分析 方法与应用

● 何晓群 编著

中国人民大学出版社

**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.

北京: 中国人民大学出版社, 1998.11

中国人民大学硕士研究生系列教材

ISBN 7-300-02817-9/F·862

I. 现…

II. 何…

III. 统计分析-研究生教育-教材

IV. C813

中国版本图书馆 CIP 数据核字 (98) 第 22300 号

中国人民大学硕士研究生系列教材

**现代统计分析方法与应用**

何晓群 编著

---

出版发行: 中国人民大学出版社

(北京海淀路 157 号 邮编 100080)

经 销: 新华书店

印 刷: 北京市丰台区印刷厂

---

开本: 850×1168 毫米 1/32 印张: 14 插页 1

1998 年 11 月第 1 版 1998 年 11 月第 1 次印刷

字数: 350 000

---

定价: 20.00 元

(图书出现印装问题, 本社负责调换)

# 中国人民大学研究生教材工作委员会

---

主任：林 岗

副主任：杜厚文 周新城 秦惠民

委员：(以姓氏笔画为序)

王 彦 王 霁 刘大椿 许征帆 吴易风

胡乃武 徐二明 高培勇 曾宪义

秘书长：高培勇（兼）

# 总 序

在中国人民大学，很早就有过组织编写研究生教材的动议。但那时研究生的课程体系尚欠规范，不仅各类课程基本上按二级学科(专业)设置，口径偏窄，同一一级学科甚至二级学科(专业)的课程，差异亦很大，而且，有些课程也与本科阶段拉不开档次，教学内容重复。所以，编写研究生教材这项直接关系到研究生培养质量的基础性工作，只在某些学科或专业做过尝试，并未提上整个学校的议事日程。

1996年4月召开的国务院学位委员会第十四次会议，是我国研究生教育的一个重大转折点。在这次会议上，明确提出了“以提高质量为中心”的研究生教育工作思路，并且根据培养跨世纪复合型人才的需要，相应调整了研究生学科专业目录。以此为契机，从1997年起，我们对硕士研究生课程体系进行了较大规模的改革。(1)适当加大课程教学在整个硕士研究生培养过程中的比重。(2)将本科、硕士、博士三个培养阶段的课程设置作为一个系统，通盘考虑，拉开档次。按照硕士研究生的培养目标设置课程，确定教学内容。(3)本着“宽口径，厚基础”的原则，拓宽硕士研究生的知识领域，在学科门类内设置所有专业通开的核心课程——学科基础课。(4)统一课程设置标准，建立一套由公共课、学科基础课、专业课和选修课等四个类别组成的规范化课程体系。

打下了较为规范化的课程体系这个基础以后，我们即着手组

织编写硕士研究生系列教材。所编教材的重点是全校所有学科通开的公共课教材以及分别在各个学科门类通开的学科基础课教材。现在呈现在读者面前的，便是我们组织编写的第一批硕士研究生教材。

编写研究生教材，本身就是一件新鲜事物，编写建立在规范化的课程体系基础上的硕士研究生公共课和学科基础课教材，难度就更大了。因此，我们迈出的这一步，既具有探索性，又具有深远的意义和影响。当然，我们深知，这一批以及今后陆续出版的硕士研究生教材，肯定存在着这样或那样的不足，还需在使用过程中不断充实、完善。我们也深信，只要方向明确，经过一个时期的努力，我校的研究生课程体系和教材建设将会日臻成熟。

**中国人民大学研究生教材工作委员会**

1998年3月22日

## 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。如金融风险、通货膨胀、税务稽查、保险精算、环境保护、持续发展、国际竞争、公众意向等，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。面对 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、财经管理类硕士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科学、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，本着提高硕士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。为重点突出方法思想和应用，每种方法尽可能紧密结合中国社会、经济、管理方面的实际问题，以案例研究为

DA370/05



导向，为非统计专业硕士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该书内容。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、任兵处长，统计学系赵彦云、易丹辉、顾岚教授，经济科学出版社莫霓舫、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。我系兼职教授、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。台湾辅仁大学统计学系主任谢邦昌教授与笔者的多次交谈，也给我许多有益启示。我系博士生孟生旺、刘勤、宋光辉、李序颖、刘长标等为本书的部分案例做过一些验算。合肥工业大学戴俭华教授，西安统计学院王国梁教授，华北电力大学张永林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

<b>第一章 概论</b> .....	1
§ 1.1 重新认识统计学 .....	1
§ 1.2 市场呼唤统计学 .....	3
§ 1.3 统计分析方法及其应用概述 .....	6
本章思考与练习 .....	16
<b>第二章 统计学基础知识回顾</b> .....	17
§ 2.1 统计数据的整理与描述.....	17
§ 2.2 随机向量的数字特征.....	22
§ 2.3 几种重要的概率分布.....	25
§ 2.4 参数估计.....	31
§ 2.5 假设检验.....	33
本章思考与练习 .....	38
<b>第三章 定性变量的统计分析</b> .....	40
§ 3.1 多项分布与 $\chi^2$ 检验 .....	40
§ 3.2 列联表分析.....	45
§ 3.3 一致性检验.....	52
§ 3.4 拟合优度检验.....	55
本章思考与练习 .....	59
<b>第四章 一元线性回归</b> .....	61
§ 4.1 一元线性回归模型.....	61

§ 4.2	参数 $\beta_0$ 、 $\beta_1$ 的最小二乘估计 .....	70
§ 4.3	回归方程的显著性检验 .....	77
§ 4.4	可化为一元线性回归的曲线回归 .....	83
§ 4.5	回归方程的应用——预测和控制 .....	90
	本章思考与练习 .....	100
<b>第五章</b>	<b>多元线性回归</b> .....	<b>102</b>
§ 5.1	多元线性回归模型 .....	102
§ 5.2	回归参数的最小二乘估计(OLSE) .....	107
§ 5.3	回归方程的显著性检验 .....	110
§ 5.4	回归系数的显著性检验 .....	114
§ 5.5	多元线性回归模型的应用 .....	123
	本章思考与练习 .....	130
<b>第六章</b>	<b>违背基本假设的回归分析</b> .....	<b>132</b>
§ 6.1	异方差性的产生背景及其处理 .....	132
§ 6.2	序列相关性的产生背景及其处理 .....	147
§ 6.3	多重共线性的产生背景及其处理 .....	163
	本章思考与练习 .....	175
<b>第七章</b>	<b>自变量选择与逐步回归</b> .....	<b>179</b>
§ 7.1	自变量选择对估计和预测的影响 .....	179
§ 7.2	关于自变量选择的几个准则 .....	185
§ 7.3	逐步回归及其应用 .....	193
§ 7.4	自变量中含有定性变量的回归模型 .....	204
	本章思考与练习 .....	212
<b>第八章</b>	<b>聚类分析</b> .....	<b>215</b>
§ 8.1	聚类分析的基本概念 .....	215
§ 8.2	距离和相似系数 .....	218
§ 8.3	系统聚类法 .....	223
§ 8.4	模糊聚类分析 .....	241

本章思考与练习·····	249
<b>第九章 判别分析</b> ·····	252
§ 9.1 距离判别·····	253
§ 9.2 费歇尔判别·····	264
§ 9.3 贝叶斯判别·····	268
本章思考与练习·····	276
<b>第十章 主成分分析</b> ·····	281
§ 10.1 主成分分析的基本思想·····	281
§ 10.2 主成分的几何意义与一般数学模型·····	283
§ 10.3 主成分的求法及性质·····	286
§ 10.4 主成分分析的计算步骤·····	293
§ 10.5 主成分分析的应用·····	305
本章思考与练习·····	315
<b>第十一章 因子分析</b> ·····	316
§ 11.1 因子分析的基本思想·····	316
§ 11.2 因子分析模型·····	317
§ 11.3 因子载荷的统计意义·····	319
§ 11.4 因子载荷矩阵的求解·····	322
§ 11.5 因子旋转·····	328
§ 11.6 因子得分·····	332
本章思考与练习·····	343
<b>第十二章 对应分析</b> ·····	345
§ 12.1 对应分析的基本思想·····	345
§ 12.2 对应分析方法的原理·····	347
§ 12.3 对应分析的计算与应用·····	353
本章思考与练习·····	370
<b>第十三章 典型相关分析</b> ·····	373
§ 13.1 典型相关分析的基本思想·····	373

§ 13.2	总体典型相关和典型变量	375
§ 13.3	样本典型相关和典型变量	378
§ 13.4	典型相关系数的显著性检验	379
§ 13.5	典型相关应用实例	380
	本章思考与练习	395
附录 A	矩阵代数	397
§ A.1	向量与矩阵	397
§ A.2	矩阵的分解和微商	408
附录 B	统计表	411
表 1	泊松分布表	411
表 2	标准正态分布表	413
表 3	正态分布分位数表	416
表 4	卡方( $\chi^2$ )分布表	419
表 5	相关系数检验表	420
表 6	$t$ 分布的临界点	421
表 7	$F$ 分布表	422
表 8	$D.W$ 检验上下界表	432
参考文献		434

# 第一章 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 重新认识统计学

人文社会科学、财经管理类研究生在大学阶段大都学过“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。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年以前中国的社会经济统计学,基本上是前苏联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”。统计学是否具有阶级性我们姑且不论,但前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影响是不容置疑的。

在中国，人们把统计学确立为前苏联统计模式后，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，把数理统计当作资产阶级货色加以批判，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圈子里。前苏联的统计模式对高度集中的计划经济起了一定作用，但这种模式也拉大了中国统计与世界统计之间的差距。

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。推断统计是在搜集、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。不难想像，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年教育部进行的专业调整也将统计学归入理学类一级学科。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。在国际惯例上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。

需要注意的是,我们现在所说的统计学与过去的社会经济统计学相比有着很大的差别。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。统计学给应用带来新思维,应用给统计学带来活力。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参考文献[1]。

## § 1.2 市场呼唤统计学

统计学在中国被列为一级学科,这只是统计与国际接轨的第一步。由于统计界长期受前苏联模式的禁锢,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。

中国目前正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,前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思辨性的思维以深厚的底蕴,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