

MBA精品课程系列教材·核心课系列

# 数据建模与决策

Master  
of  
Business  
Administration

刘国山 等 编著

 中国人民大学出版社

MBA精品课程系列教材·核心课系列

# 数据建模与决策



刘国山 等 编著

 中国人民大学出版社

## 图书在版编目 (CIP) 数据

数据建模与决策/刘国山等编著.  
北京:中国人民大学出版社, 2004  
(MBA 精品课程系列教材)

ISBN 7-300-05861-2/F·1869

- I. 数…
- II. 刘…
- III. 数据库技术-建立模型-研究生-教材
- IV. TP311.13

中国版本图书馆 CIP 数据核字 (2004) 第 084724 号

MBA 精品课程系列教材·核心课系列

**数据建模与决策**

刘国山 等 编著

---

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) 010-62511239 (出版部)  
010-82501766 (邮购部) 010-62514148 (门市部)  
010-62515195 (发行公司) 010-62515275 (盗版举报)

网 址 <http://www.crup.com.cn>  
<http://www.ttrnet.com> (人大教研网)

经 销 新华书店

印 刷 河北涿州星河印刷有限公司

开 本 787×1092 毫米 1/16 版 次 2004 年 10 月第 1 版  
印 张 30 插页 2 印 次 2004 年 10 月第 1 次印刷  
字 数 533 000 定 价 36.00 元

---

版权所有 侵权必究 印装差错 负责调换



## 内容简介

本书以MBA学生的数学基础为起点，同时注意深入浅出，系统地介绍常用的数量化模型。我们以介绍统计模型为主，适当地介绍了一些管理科学方面的模型。本书的每一章以案例为先导，然后介绍相关的知识，最后以一般的练习来巩固知识，并在问题的解决中强调计算机软件的利用和写作方式。鉴于Excel软件应用比较普及，而且操作简单，比较容易学习，在本书中主要以Excel作为工具来解决教材中几乎所有的计算问题。

本书适合MBA学生以及经济和管理方向的硕士研究生使用。

# 总 序

随着经济全球化进程的不断发展，以 MBA 教育为代表的中国管理教育日益受到关注，而 MBA 教育经过 13 年的风雨历程，也取得了长足的发展，引发了人们更加深刻地思考：如何进一步提升中国 MBA 教育的教学质量？

首先，经济全球化对国际化管理人才的素质提出了更高的要求。经济的全球化带来的产业升级和企业的国际化战略将是未来企业求得生存与发展的关键，越来越复杂的经济环境使跨国公司对具有全球性眼光的国际型人才需求日趋增多。经济全球化时代要求企业经营者彻底转变经营思想和经营理念，具备竞争观念、创新意识和开拓精神，具备全球性的敏锐洞察力，通晓多国语言，有雄厚的技术功底和创业精神、有能力管理日益细分和分散的组织。

其次，随着中国市场经济的日益完善，数以万计的民营企业迫切需要自己的企业家和职业经理人，如何使 MBA 教育本土化更是一个迫在眉睫的问题。如果说前 13 年 MBA 教育的发展更多走的是拿来主义的道路，那么现在更需要探索一条本土化的道路。

最后，随着中国企业改革的深化，专业 MBA 教育的时代来临了。现在不仅出现了 MPA、MPAcc 等专业硕士学位，在 MBA 中也向专业细分方向发展；不仅出现了市场营销方向、人力资源管理等专业方向的细分课程，而且出现了涉农管理 MBA、航空管理 MBA、电力资源管理 MBA 等等。正是这种专业方向的 MBA 教育，为 MBA 教育提供了更加广阔的市场。

在这样的市场需求下，编写一套国际化与本土化相结合、更加体现专业和行业细分的教材势在必行。中国人民大学出版社在全国 MBA 专业指导委员会的指导下，组织国内外一流专家，编写了“MBA 精品课程系列教材”。这套教材将很好地适应 MBA 教育的发展，也使 MBA 的毕业生更好地适应用人单位的需要，最终提升中国企业的管理水平，促进中国经济的进一步发展。

# 出 版 说 明

中国人民大学出版社将陆续推出“MBA 精品课程系列教材”、“管理学研究生参考书系”，以满足规模越来越大的研究生层次的教学需要。

目前社会上已经引进了一批国外的 MBA 教材，但随着中国经济改革理论与实践的飞速发展，人们渴望学习包含更多中国企业实战经验的教材。我们将聘请国内各领域的一流专家，他们既熟悉西方经济管理理论和教学模式，又熟悉中国企业现实，有较为丰富的 MBA 教学经验。因此“MBA 精品课程系列教材”将是一套本土化的、具有较强实战性的教材。这套教材所涵盖的课程，不仅包括核心课，而且包括人力资源、市场营销、会计学、金融与财务管理、运营管理等专业方向课。我们希望通过不同学科模块的教学，使 MBA 教育更具针对性，也使 MBA 毕业生的专业素质和能力更高。在形式上，这套教材既包括传统的纸质教材，又包括案例和网上教学资源，这将更加方便教师教学。

“管理学研究生参考书系”的读者对象则主要为管理学硕士研究生和博士研究生。这套丛书将以出版国内外管理学的前沿著作和经典文献为主。目前我们已精选了部分国外优秀作品，如日本有斐阁出版的《商品流通》、布莱克威尔出版社出版的《组织行为学原理手册》、世界科技出版公司出版的《企业家精神》等。我们将陆续推出更多的优秀作品，同时也期待着广大读者向我们积极推荐。

愿这两套丛书能为我国 MBA 和管理学研究生教育事业的发展做出应有的贡献。

中国人民大学出版社

2004 年 9 月

# 前 言

“数据建模与决策”是全国 MBA 教育指导委员会指定的 MBA 核心课程，因此，这门课程的重要性是不言而喻的。但是，到目前为止，我国还没有一本合适的教材。据我们所知，到目前为止，仅有两本关于数据建模与决策的书，而这两本书也只包括管理科学的内容而不包括统计的内容。也就是说，虽然这两本书的书名叫《数据建模与决策》，但内容却与我们要编写的《数据建模与决策》有很大的差别。

编写这门课程所使用的教材，遇到了编写许多其他教材所没有遇到的困难。对于“数据建模与决策”课程应包含的主要内容，国内开设 MBA 课程的院校以及全国 MBA 教育指导委员会的意见基本上是一致的，即应包括管理科学（运筹学）和应用（商务）统计两门课程的主要内容。但是在具体内容的选择上，每个学校就有非常大的差异，因为要讲述的内容太多，必须作一些取舍。如果同时考虑到读 MBA 的学员的数学基础，那么完成这门课程的教学任务就是非常困难的。这就导致了《数据建模与决策》教材编写中对于内容如何进行取舍的问题。

“数据建模与决策”课程的教学目的是以教授应用技能为主。由于实际问题异常复杂，几乎所有数据的处理和模型的解决都需要借助计算机来实现，因此，选择什么样的软件来处理就是一个问题。如果选择非常专业的软件，则要花费许多时间学习如何使用这些软件，从而影响教学的其他环节。但是，使用比较方便的软件一般又不能有效地解决问题。而且，如果统计部分和管理科学部分使用不同的软件，那么学生就需要在短时间内掌握两个或两个以上的软件的操作。

虽然全国 MBA 教育指导委员会在对于 MBA 教学检查的要求中对“数据建模与决策”课程的教学案例未做要求，但是，各个学校及其 MBA 学生对于案例教学的热中使得我们在教学中不得不重视案例教学问题。这给本来时间压力就比较大的“数据建模与决策”课程的教学任务又增添了新的困难。这也是我们编写《数据建模与决策》教材所必须面对的问题。

基于编写《数据建模与决策》教材所面临的问题，编写者们经过集体协商，决定以中

国人民大学 MBA 的“数据建模与决策”课程的基本教学内容为主，同时参考大连理工大学管理学院的教学内容，来编写可供一个学期的课程所使用的教材。在内容选取上，我们以介绍统计内容为主，适当地介绍一些管理科学方面的模型。这样做的主要原因是教学时数的限制，在有限的时间内是很难完成太多的教学任务的。

我们参照国外的 MBA 教材，确定了以案例为先导，然后介绍相关知识，最后以一般的练习来巩固知识，并在解决问题的过程中强调计算机软件的利用的写作方式。至于案例作业，我们没有编写，其主要原因是现在有许多的案例教材可以选用。

由于在实际应用中数量分析方法所涉及的数据的规模一般都很大，需要通过相应的计算机软件来实现，因此经过商议，在向国内外同行进行咨询的基础上，我们决定在《数据建模与决策》教材中主要以 Excel 作为工具来解决教材中几乎所有的计算问题，包括管理科学中的计算和统计数据的处理。这样做的原因是 Excel 软件应用比较普及，几乎在任何一台电脑上都可以使用 Excel。而且，Excel 操作简单，比较容易学习，不会占用太多的教学时间。同时，电子表格也是现在企业管理中一个非常有用的工具。

本书第 1 章由深圳大学管理学院李丽教授编写，第 2 章至第 11 章由中国人民大学统计系的贾俊平教授编写，第 12 章和第 13 章由中国人民大学商学院的刘国山教授编写，第 14 章和第 15 章由大连理工大学管理学院的胡祥培教授编写。邓茶亚、陈守龙和谭英平也参加了编写工作。

由于时间所限，编写错误在所难免，希望各位同行和读者不吝赐教，我们将在广泛征求 MBA 学生意见和建议的基础上作进一步的修改。

作者

2004 年 2 月



# 目 录

## 第 1 章

概 述 .....	(1)
第 1 节 数据的类型 .....	(2)
第 2 节 数据的来源 .....	(3)
第 3 节 数据与决策 .....	(7)
第 4 节 模型的类型 .....	(9)
第 5 节 模型与决策 .....	(12)
小 结 .....	(13)
练 习 题 .....	(13)

## 第 2 章

数据的整理与显示 .....	(15)
第 1 节 数据的预处理 .....	(17)
第 2 节 分类和顺序数据的整理与显示 .....	(21)
第 3 节 数值型数据的整理与显示 .....	(27)
第 4 节 统计表 .....	(41)
小 结 .....	(42)
练 习 题 .....	(43)

## 第 3 章

数据分布特征的测度 .....	(48)
第 1 节 集中趋势的测度 .....	(49)
第 2 节 离散程度的测度 .....	(59)

第3节 偏态与峰态的测度 .....	(68)
小结 .....	(72)
练习题 .....	(72)

**第4章**

概 率 .....	(76)
-----------	------

第1节 事件及其概率 .....	(77)
第2节 离散型概率分布 .....	(88)
第3节 连续型概率分布 .....	(100)
小结 .....	(113)
练习题 .....	(114)

**第5章**

抽样与抽样分布 .....	(117)
---------------	-------

第1节 概率抽样方法 .....	(118)
第2节 三种不同性质的分布 .....	(120)
第3节 一个总体参数推断时样本统计量的抽样分布 .....	(121)
第4节 两个总体参数推断时样本统计量的抽样分布 .....	(129)
小结 .....	(132)
练习题 .....	(133)

**第6章**

参数估计 .....	(135)
------------	-------

第1节 参数估计的一般问题 .....	(137)
第2节 一个总体参数的区间估计 .....	(143)
第3节 两个总体参数的区间估计 .....	(149)
第4节 样本容量的确定 .....	(159)
小结 .....	(162)
练习题 .....	(163)

**第7章**

假设检验 .....	(167)
------------	-------

第1节 假设检验的基本问题 .....	(169)
---------------------	-------

第2节 一个总体参数的假设检验 .....	(176)
第3节 两个总体参数的假设检验 .....	(192)
小结 .....	(209)
练习题 .....	(210)

**第8章**

方差分析与试验设计 .....	(214)
-----------------	-------

第1节 方差分析引论 .....	(216)
第2节 单因素方差分析 .....	(221)
第3节 方差分析中的多重比较 .....	(230)
第4节 双因素方差分析 .....	(232)
第5节 试验设计初步 .....	(242)
小结 .....	(247)
练习题 .....	(247)

**第9章**

一元线性回归 .....	(250)
--------------	-------

第1节 变量间关系的度量 .....	(251)
第2节 一元线性回归 .....	(258)
第3节 利用回归方程进行估计和预测 .....	(271)
第4节 残差分析 .....	(274)
小结 .....	(280)
练习题 .....	(281)

**第10章**

多元线性回归 .....	(286)
--------------	-------

第1节 多元线性回归模型 .....	(288)
第2节 回归方程的拟合优度 .....	(291)
第3节 显著性检验 .....	(293)
第4节 多重共线性 .....	(296)
第5节 利用回归方程进行估计和预测 .....	(300)
第6节 虚拟自变量的回归 .....	(301)
第7节 非线性回归 .....	(310)

小结 .....	(313)
练习题 .....	(314)

**第 11 章**

<b>时间序列分析和预测</b> .....	(319)
------------------------	-------

第 1 节 时间序列及其分解 .....	(320)
第 2 节 时间序列的描述性分析 .....	(322)
第 3 节 平稳序列的平滑和预测 .....	(325)
第 4 节 有趋势序列的分析和预测 .....	(331)
第 5 节 复合型序列的分解 .....	(343)
小结 .....	(349)
练习题 .....	(350)

**第 12 章**

<b>线性规划</b> .....	(355)
-------------------	-------

第 1 节 线性规划的模型 .....	(356)
第 2 节 线性规划的性质 .....	(364)
第 3 节 线性规划的计算机解法 .....	(367)
第 4 节 线性规划模型的应用 .....	(377)
小结 .....	(387)
练习题 .....	(388)

**第 13 章**

<b>运输问题和指派问题</b> .....	(397)
------------------------	-------

第 1 节 运输问题 .....	(398)
第 2 节 运输问题的应用 .....	(399)
第 3 节 指派问题 .....	(408)
第 4 节 指派问题的应用 .....	(412)
小结 .....	(418)
练习题 .....	(418)

**第 14 章****动态规划** ..... (423)

第 1 节 多阶段决策问题 ..... (424)

第 2 节 动态规划问题的逆序算法 ..... (427)

第 3 节 动态规划的基本概念及最优化原理 ..... (430)

第 4 节 离散确定型动态规划模型的求解 ..... (434)

第 5 节 连续确定型动态规划模型的求解 ..... (440)

小结 ..... (442)

练习题 ..... (442)

**第 15 章****图与网络分析** ..... (446)

第 1 节 图的基本概念与模型 ..... (447)

第 2 节 树图和图的最小部分树 ..... (449)

第 3 节 最短路问题 ..... (454)

第 4 节 中国邮路问题 ..... (459)

第 5 节 网络的最大流 ..... (460)

小结 ..... (464)

练习题 ..... (465)

# 第 1 章

## 概 述

### 本章要点

- 数据的类型
- 数据的来源
- 数据与决策
- 模型与决策

### ABC 公司的决策

ABC 公司是一家生产特种钟表的企业，它向用户提供能满足其特殊要求的产品。新产品投入市场的过程为：首先公司的研发部开发新产品，然后公司决策层决定是否投入批量生产，而不是采用订单销售的模式。在上个月，研发部设计了一种新款手表，公司决策层例会讨论是否生产这种手表，如果生产要确定生产多少。会上争论很激烈。有人说，这要看市场是否有需求，需求有多少。还有人说要根据这种产品的生产成本来确定是否生产。表面看来这两种说法都没错，但问题是市场需求数据从哪来？成本如何估计？另外，这些数据都具有一定的弹性与不确定性，那么如何来判断该不该生产，生产多少呢？一时间难以抉择。

事实上，世界上每天都有许许多多类似的抉择摆在管理者面前，他们有的做出了正确的判断，有的却造成很大失误。为什么会出现这种差异呢？有没有一种工具能较为有效地帮助他们做出正确的决策呢？答案是肯定的，这就是数量决策。

数量决策，顾名思义是根据所要解决的问题的相关数字信息，或描述其规律的数学模型，分析解决疑惑，使管理者在决策时能够有据可依，从而做出更理性、更科学的决策。

本章将重点介绍一些常用名词术语，解释数据、模型和决策的基本概念，分析三者之间的联系，同时将简单介绍一下相关的应用软件，引导大家进入一个量化而又生动的空间。

## 第 1 节 数据的类型

我们每天都会遇到很多的数字和信息。那么到底什么是数据呢？按统计学的定义，数据是采用某种计量尺度对事物进行计量的结果。也就是说，它是事物某种特征性质的反映。事实上，对不同的事物，我们能够予以计量或者说测量的程度是不同的，而且事物的属性不同必然会导致数据类型的不同。比如人的性别和人的学历就不可比，物体的温度和物体的重量也不可比。通常，按计量学的分类方法，按照对事物计量的精确程度，可将我们所用的计量尺度由粗到细、由低级到高级分为四个层次，即定类尺度、定序尺度、定距尺度和定比尺度。

定类尺度是最粗略、最低层次的计量尺度，它按事物的某种属性对其简单的分类。如性别分为男女，颜色分为七色。

定序尺度是对事物之间等级差别或顺序差别的一种测量。它不仅可按属性分类，还可确定该事物的优先级或顺序。例如，一个人不仅能分出性别，还能分出他的职位高低、学历高低。这种尺度比定类尺度高级，进一步描述了事物的属性，但还没有达到能用数字精确确定事物属性的地步，没有数学运算的功能，因而对事物的描述具有一定的局限性。

定距尺度是将事物属性精确地划分，明确地指出事物的不同。就拿物体的温度来说，它在分清物体的类别和大小后，精确地用数字表示了物体间特性的差异。20℃和30℃的差异是非常清楚的。它可运用数学运算法则来表示两者间相差10℃。

定比尺度比定距尺度的属性更高一个层次，它用比值、比率来描述事物的差别。比如北京大学校园的面积比清华大学校园的面积小，用定距尺度表示就是二者相差1 248亩，用定比尺度表示就是二者相差0.76倍。

不同的计量尺度会得到不同类型的统计数据。根据计量尺度计量的结果，我们将数据相应地分为以下四种类型：定类数据（只能区分物体的类别，不能区分顺序，比如民族、材质等），定序数据（除了表示类别外，还可描述顺序、等级，如产品质量等级、温度高低等），定距数据（主要以数字形式表现，可以进行加、减运算，如学生考试成绩、物体长度等），定比数据（主要以数字形式表现，可

进行加、减、乘、除运算，常表现成比率形式，如工人工资、体重等)。前两类数据描述的是事物的属性特征，无法用具体明确的数值表示，只能对事物的性质进行表征，所以常被人们称为定性数据；后两类数据说明事物的数量特征，能够用明确的数值来表现，所以常被称为定量数据或数量数据。

为什么要区分这些不同类型的数据呢？因为对不同类型的数据我们需要采用不同的统计方法来处理和分析。例如，对定类数据，通常计算出各组的频数或频率，计算其众数和异众比率，进行列表分析和检验等；对定序数据，可以计算其中位数和四分位差，计算等级相关系数等；对定距或定比数据还可以用更多的统计方法进行处理，如计算各种统计量、进行参数估计和检验等。我们处理的大多为数量数据。

在本节的最后，我们引入几个后面常用的概念。

**总体：**指被分析事物的个体的集合，也就是样本收集者感兴趣的个体或对象的完全集合。例如，全体人类的集合是一个总体。在统计中，总体可能是全体动物的集合，或是某种产品的集合，或是加州全体红杉树的集合。一般说来，直接对总体进行研究是比较困难的，因此，习惯的做法是在总体中选择一部分个体，也就是样本来研究。

**样本：**是总体的一个子集。样本由样本收集者从总体中选出的个体、对象或测量值组成。

**参数：**是描述总体所有数据的数字特征的数值。例如，某一所学院的学生入学“平均”年龄是一个总体参数。参数是描述整个总体的一个特征值，通常用希腊字母表示。当我们研究特定的参数时，将会规定这些参数相应的记号。

**统计量：**是概括样本数据的一个数值。对于每一个参数，有一个相应的样本统计量。用25个学生身高数据求得的“平均”身高，是样本统计量。统计量是描述样本特征的一个值。

## 第2节 数据的来源

要利用统计方法帮助我们分析和解决问题，首先必须做好资料 and 信息的收集工作，这其中需要收集大量的数据。获得数据有多种方式，一般可分为直接方式和间接方式。



## 数据的直接来源

收集数据资料的直接方式是指通过亲身调查得来数据，即第一手资料，通常的方法有普查、随机抽样调查、统计报表、选择样本调查等方式。

**普查：**是为收集准确信息，由专人组织的一次性全面调查。它应用于需要掌握某一项事物的全面信息的调查工作，是获取全面资料的有效途径。普查起初主要用于人口统计，后来被大量用于许多社会问题的调查，如国家经济问题、GDP统计等。虽然普查能提供翔实准确的信息，但每一次普查活动，会涉及各个方面的单位、企业和个人，从普查设计、培训、实施到结果分析报告，都需要耗费大量的人力、物力和财力。因此，它并不适合经常使用，也不是任何情况下都适用的。

**随机抽样调查：**是从总体中抽取一定量的样本进行分析研究，以确定总体的性质特点。随机抽样调查具有经济、代表性强、较准确等特点，它既具备一般非全面调查的长处，同时又可以通过科学推算达到对总体的认识，因而是目前最流行的数据调查手段。以社会经济管理领域为例，经常采用随机抽样调查收集资料的有家电市场调查、民意测验、工农业生产调查、抽样审计等。

**选择样本调查：**是从总体中选择具有代表性的典型对象进行研究分析，借以推断总体的本质和变化规律。样本调查中的样本，是刻意挑选出来的，它比随机抽样调查更经济、针对性更强。这种调查方式类似于通常所说的“典型调查”。

**统计报表：**是指一个部门内部根据有关规定，由上而下地布置，自下而上地逐级上报数据的一种调查方式。

**观察和试验：**是指调查者通过眼睛或借助仪器来了解所调查对象的有关信息。观察主要是旁观，并不触及被调查对象，比如站在天桥上数路上每分钟过往的车辆数，或在商场门口记录每天的客流量。这种方式需要调查者具有一定的经验和敏捷的感官。试验就是调查者通过设计，对被调查对象进行实验而得到相关资料的方式。比如同样想了解车流量，实验者也可以通过设置红灯，记录红灯后压车的数量等手段达到目的。从收集资料的角度看，试验设计与观察调查的区别主要表现在：观察调查资料是“自然资料”，是调查对象业已存在的、自然而然发生的结果；而试验资料是“生成资料”，是人为控制试验条件产生的。在收集资料过程中，资料收集者的地位不同，观察调查中资料收集者是被动的，只能进行观察和记录，不能试图改变什么；而在试验设计中资料收集者却是主动的，可以在一定范围内自由搭配试验因素。注意到试验资料与调查资料之间的这种区别很有必要，因为这将涉及对资料处理方法的选择，以及统计解释问题。