

21世纪MBA 规划教材



CD-ROM

数据、模型与决策

基于Excel的建模和商务应用 (第二版)

DATA, MODELS AND DECISIONS

蒋绍忠 编著



北京大学出版社
PEKING UNIVERSITY PRESS

数据、模型与决策

基于Excel的建模和商务应用 (第二版)

DATA, MODELS AND DECISIONS

蒋绍忠 编著



TP391.13-43
05-2
P



北京大学出版社
PEKING UNIVERSITY PRESS

图书在版编目(CIP)数据

数据、模型与决策:基于 Excel 的建模和商务应用/蒋绍忠编著.—2 版.—北京:北京大学出版社,2013.4

(21 世纪 MBA 规划教材)

ISBN 978 - 7 - 301 - 22263 - 8

I. ①数… II. ①蒋… III. ①表处理软件 - 应用 - 商业管理 - 工商管理硕士 - 教材
IV. ①F716 - 39

中国版本图书馆 CIP 数据核字(2013)第 042535 号

书 名: 数据、模型与决策——基于 Excel 的建模和商务应用(第二版)

著作责任者: 蒋绍忠 编著

责任编辑: 李 娟

标准书号: ISBN 978 - 7 - 301 - 22263 - 8/F · 3578

出版发行: 北京大学出版社

地 址: 北京市海淀区成府路 205 号 100871

网 址: <http://www.pup.cn>

电子信箱: em@pup.cn QQ:552063295

新浪微博: @北京大学出版社 @北京大学出版社经管图书

电 话: 邮购部 62752015 发行部 62750672 编辑部 62752926 出版部 62754962

印 刷 者: 北京世知印务有限公司

经 销 者: 新华书店

787 毫米 × 1092 毫米 16 开本 31.75 印张 791 千字

2010 年 1 月第 1 版

2013 年 4 月第 2 版 2013 年 4 月第 1 次印刷

印 数: 0001—3000 册

定 价: 65.00 元(附光盘)

未经许可,不得以任何方式复制或抄袭本书之部分或全部内容。

版权所有,侵权必究

举报电话:010-62752024 电子信箱:fd@pup.pku.edu.cn

教师反馈及教辅申请表

北京大学出版社以“教材优先、学术为本、创建一流”为目标，主要为广大高等院校师生服务。为更有针对性地为广大教师服务，提升教学质量，在您确认将本书作为指定教材后，请您填好以下表格并经系主任签字盖章后寄回，我们将免费向您提供相应教辅资料。

书号/书名/作者			
您的姓名			
校/院/系			
您所讲授的课程名称			
每学期学生人数	_____ 人 _____ 年级	学时	
您准备何时用此书授课			
您的联系地址			
邮政编码		联系电话 (必填)	
E-mail (必填)		QQ	
您对本书的建议:			系主任签字 盖章

我们的联系方式:

北京大学出版社经济与管理图书事业部

北京市海淀区成府路 205 号, 100871

联系人: 徐冰

电话: 010-62767312 / 62757146

传真: 010-62556201

电子邮件: em@pup.cn xubingjn@yahoo.com.cn

网址: <http://www.pup.cn>

微博: @北大出版社经管图书, <http://weibo.com/pupem>

第二版前言

本书自第一版出版以来,不少读者对本书的内容和结构提出了很多宝贵的建议,指出了其中的一些错误和欠妥之处。对于大家的指导和帮助,本人深表感谢。通过和大家的交流,作者感到广大读者,特别是 MBA 任课教师和 MBA 学生,对本书的内容还是肯定和欢迎的,作者也感受到大家对本书进一步改版的期待。

作者从一年前就开始为改版做准备。2011 年秋冬季浙江大学管理学院 MBA 教学中,试用了改版以后的 PPT 演示文稿和其他教辅材料,学生反映良好。同时,在教学过程中,形成了与新内容配套的例题、习题、习题解答以及新的课堂视频文档。在这些工作的基础上,最后完成了本书的第二版书稿。

第二版主要有以下一些改动:

1. 数据处理工具改为 Excel 2007。Excel 2007 无论是功能、界面还是稳定性都比 Excel 2003 有明显的改进,已经逐渐成为主流版本。很多原来熟悉 Excel 2003 的学生,开始使用 Excel 2007 时,都觉得两者界面变化太大,许多原来熟悉的功能不知去向,有点无所适从。经过一个学期的学习,大家都觉得这样的改变是很值得的。Excel 2010 和 Excel 2013 的功能与界面和 Excel 2007 非常接近,本书中的模型、方法、操作和文档,同样适用于 Excel 2010 和 Excel 2013。

2. 增加了第五章“问卷调查”,内容除了问卷调查的基本概念和方法以外,着重介绍了前四章的数据分析和统计工具在问卷数据分析中的应用。对于 MBA 学生来说,问卷调查的内容无论对于他们的实际工作还是对于他们的学位论文写作,都是非常重要的。

3. 为了简化课程中一些比较复杂但很常用的计算操作,作者编制了一些 Excel 自定义函数。这些自定义函数是:个人所得税计算函数、问卷数据的克隆巴赫 α 信度系数计算函数、层次分析法判断矩阵特征向量和特征根计算函数。这些自定义函数和课程内容密切相关,使用简便灵活,既是很好的教学手段,又是有价值的实用工具。

和第一版一样,第二版有一张学习光盘随书发行,其中有与第二版内容配套的例题和习题的 Excel 2007 文档以及以上自定义函数的加载宏文件。另外,作者还准备了与第二版书稿配套的 PPT 教学演示文档、习题解答、课堂视频等教师用教辅材料。教师如果采用本书作为教材,只要向出版社递交书后的“教师反馈及课件申请表”,就可以从出版社免费获取第二版的教师用教辅材料。

在完成第二版书稿之际,作者再次对关心和支持本书的老师、学生、读者以及北京大学出版社表示由衷的感谢,并继续期待得到各位的指教。作者的电子邮箱仍然是:jiangsz@zju.edu.cn。

蒋绍忠

2012年10月

于浙江大学紫金港校区

第一版前言

无论是教师、医生、公司职工还是公务员,工作或多或少与数据处理有关。学生成绩的统计、治疗效果的评估、产品销量的预测、客户资料的分类、人事档案的管理等,都涉及数据处理。随着计算机以及应用软件的普及,人人都具备数据分析和处理的条件,可是并非人人都能够科学有效地分析和处理数据。

有了数据,还需要借助理论和模型,才能获得对决策有用的信息。数据是基础,模型是工具,决策是目的。为了用数据辅助决策,还需要学习一些相关的理论和模型,这些理论和模型包括概率、统计、预测、模拟和优化等,还需要掌握相应的计算机软件工具。本书就是从理论模型和软件工具两个方面,比较全面地介绍数据分析和处理的方法和技能。本书的对象是那些有数据分析处理需求,具备计算机和软件操作基础,希望进一步学习数据处理的理论、模型、技术和方法的读者。

本书有以下几个特点:

(1) 注重理论和方法相结合。书中有关的理论,都有相应的实现方法举例,其中绝大多数例子,都有软件操作界面的屏幕截图,并且对截图进行详细的标注。

(2) 介绍最常用的数据分析模型,如回归模型、预测模型、模拟模型、优化模型和层次分析模型。模型是对数据进行深层次分析的必要工具,掌握这些模型对提高数据分析和处理能力以及将数据分析和处理与决策结合起来,是非常重要的。

(3) 数据分析和处理软件采用 Excel。众所周知,Excel 是应用最广泛的电子表格,具有强大和完善的数据处理功能,Excel 表格和图形与 Word 等其他 Office 工具完全兼容。对于非专业化的数据处理人员,以 Excel 作为数据分析和处理工具,具有起点低、见效快的优势,显然是一个最佳的选择。

(4) 学习数据分析和处理,操作练习是十分必要的。为此,本书附有一张学习光盘,其中有本书所有例子和习题的 Excel 文档,以方便读者在阅读本书

时,参照这些文档进行操作练习。

(5) 编写本书的直接目的,是为工商管理硕士(MBA)学生提供一本数据、模型与决策课程的教材。为此,本书还有一张配套的教师光盘,其中包括书中所有习题的解答、长达900多页的PPT演示文档,以及长达几十个小时的课堂教学视频文档。教师光盘只赠送给选用本书作为教材的老师。

本书是作者多年从事数据、模型与决策课程教学的成果。尽管在编写和审阅中尽了最大努力,但限于作者的水平、时间和精力,深知其中错误和疏漏在所难免,恳请读者指正并告知。作者的电子邮箱是:jiangsz@zju.edu.cn,恭候各位赐教。

蒋绍忠

2009年清明节

于求是村

目 录

第 1 章 数据和数据展示	(1)
1.1 数据	(3)
1.2 Excel 基础	(8)
1.3 数据透视表和数据表	(25)
1.4 数据展示图	(35)
1.5 数据的分布属性	(46)
习题	(64)
第 2 章 概率论和统计学基础	(67)
2.1 随机事件、随机变量和频数	(69)
2.2 离散型随机变量及概率分布	(73)
2.3 连续型随机变量及概率分布	(80)
2.4 统计学概述	(99)
习题	(100)
第 3 章 抽样和估计	(103)
3.1 统计抽样	(105)
3.2 样本数据的统计分析	(109)
3.3 点估计	(111)
3.4 区间估计和置信区间	(113)
3.5 置信水平、置信区间宽度和样本数	(123)
习题	(125)
第 4 章 假设检验	(127)
4.1 假设检验的基本原理	(129)
4.2 单样本假设检验	(137)
4.3 两样本假设检验	(142)
4.4 多样本假设检验方差分析	(148)
习题	(161)
第 5 章 问卷调查	(165)
5.1 问卷和问卷调查	(167)
5.2 问卷设计	(170)
5.3 量表设计	(171)

5.4	问卷的效度和信度	(175)
5.5	问卷调查案例——网上购物的问卷调查	(177)
5.6	电子问卷制作	(201)
	习题	(217)
第6章	回归分析	(219)
6.1	简单线性回归	(221)
6.2	一元线性回归	(225)
6.3	多元线性回归	(236)
6.4	非线性回归	(244)
6.5	具有属性变量的回归	(253)
6.6	正确应用回归分析方法	(256)
	习题	(257)
第7章	统计预测	(259)
7.1	预测概述	(261)
7.2	用回归方程预测	(262)
7.3	时间序列预测	(275)
7.4	预测工具 CB Predictor	(294)
	习题	(300)
第8章	风险决策和蒙特卡罗模拟	(303)
8.1	决策概述	(305)
8.2	风险决策	(306)
8.3	风险决策的效用理论	(314)
8.4	蒙特卡罗模拟	(319)
8.5	投资决策的蒙特卡罗模拟	(327)
	习题	(331)
第9章	风险分析工具 Crystal Ball	(335)
9.1	Crystal Ball 基础	(337)
9.2	Crystal Ball 在企业经营风险分析中的应用	(365)
9.3	Crystal Ball 在项目管理风险分析中的应用	(374)
	习题	(394)
第10章	管理系统模拟	(397)
10.1	模型和模拟	(399)
10.2	库存系统	(401)
10.3	库存系统模拟	(406)
10.4	排队系统的模拟	(414)
10.5	单服务台单队列排队系统模拟	(424)
	习题	(429)

第 11 章 管理系统优化	(431)
11.1 线性规划优化数学模型	(433)
11.2 线性规划问题的基本概念	(440)
11.3 用规划求解工具求解线性规划问题	(443)
11.4 线性规划问题求解结果的分析	(455)
习题	(463)
第 12 章 多目标决策	(465)
12.1 多目标决策概述	(467)
12.2 层次分析法	(469)
12.3 住宅选择的层次分析模型	(478)
12.4 层次分析法的 Excel 模型	(485)
习题	(489)
附 录	(491)
Excel 自定义函数加载宏的安装	(493)
参考文献	(497)

第 1 章 数据和数据展示

1.1 数据

1.1.1 数据概述

“气象台今天 18 点发布的报告,本市今天晴,最高气温 12.3℃,最低气温 5.1℃,偏北风 5 级。明天天气预报:小雨,10℃—6℃,微风,降水概率 70%。”

“16 日上证综指突破了 3 000 点关口,以 3 018.18 点大幅高开,午后更创下 3 036.35 点的历史新高。当日上证综指报收于 2 998.47 点,较前一交易日微涨 0.18%。”

“截至第四节 7 分钟,姚明在场上 15 投 5 中得到 16 分,8 个篮板、1 次助攻、1 次断球、1 次封盖,并有 2 次犯规和 2 次失误。”

“今天本市空气污染指数为 78,空气质量描述为良,主要污染物为可吸入颗粒物。预计明天空气污染指数为 40—50,空气质量描述为优。”

以上四段文字分别用不同类型的数据发布了不同的信息。其中,“晴”、“小雨”、“最低气温 12.3℃”、“降水概率 70%”、“上证指数 3 018.18 点”、“得分 16 分”、“8 个篮板”、“空气污染指数 78”、“空气质量描述‘优’、‘良’”等都是数据。

数据是信息的重要组成部分,每个人时时刻刻都要和数据打交道,数据成为现代社会生活不可或缺的因素。

数据根据其来源和用途可以大致分为科学数据、社会数据和商业数据。科学数据是自然现象、科学试验和工程项目的过程和结果的记录;社会数据是社会非商业活动的事件记录,如人口、教育、犯罪等数据;商业数据是经济和商业活动的记录,如宏观经济数据、股市数据、企业经营数据等。这些分类不是绝对的,例如,环境数据既是科学数据,又是社会数据。

在商业活动中,无论是商业情报、商业计划、商业报表、商业决策,都离不开数据。数据展示和处理技术是商业活动的一项基本技能。本书的核心,就是介绍商业活动中基本的数据处理方法,用各种模型分析和处理数据,以及利用数据进行科学正确的商业决策。

科学数据的主要特征是数据来源、记录过程和数据处理的客观性。科学数据要求数据的客观性和可重复性,数据的获取和处理过程应尽可能避免人为因素的影响。而大多数商业数据本身就是人为活动的结果,而且一些商业数据只能依靠问卷调查来获得。因此,和科学数据相比,商业数据中的人为因素是不可避免的。但是这并不意味着商业数据可以随心所欲地编造或篡改。和科学数据一样,数据的客观性和可重复性原则同样是商业数据获取和处理的基本原则。即在相同环境下,对于同一个问题,由不同的人一次或多次获取的数据应该没有本质的

区别。只有遵从数据客观性和可重复性原则来获取和处理数据,数据处理的结果才有意义。

本书介绍的数据处理、建模技术和决策分析方法,主要针对商业领域。对政府、教育、科研、医疗卫生等非商业行业也是适用的。

1.1.2 数据的分类

1. 数值型和属性型数据

数据可以分为数值型数据和属性型数据。数值型数据是用数值来表示的,“12.3℃”、“3018.18点”、“0.18%”、“16分”、“8个篮板”等都是数值型数据。属性型数据是用文字、判断等表示的数据,“晴”、“小雨”、“微风”、“优”、“良”等都是属性型数据。

数值型数据又可以分为连续数值型数据和离散数值型数据。“12.3℃”、“3018.18点”、“0.18%”就是连续数值型数据,“16分”、“8个篮板”等都是离散数值型数据。连续数据可以连续变化,离散数据只能取若干分散的数值,通常离散数据都是整数。

2. 静态数据和动态数据

根据数据是否随时间变化,可以分为静态数据和动态数据。静态数据不变化或在所关注的时间段内很少变化,动态数据在该时间段内随时间有明显的变化。例如,某一个人的相关数据如表 1.1 所示。

表 1.1 静态数据和动态数据

项目	数据	数据性质
姓名	王大明	静态
性别	男	静态
出生年月	1970年4月	静态
年龄	37岁	动态
出生地	北京	静态
学历	本科	静态
身高	176厘米	静态
体重	73千克	动态
年收入	13万元	动态

3. 时间序列数据、截面数据和面板数据

根据数据的时间属性,可以分为时间序列数据(简称时序数据)、截面数据和面板数据。时序数据展现不同时间段数据的变化,截面数据表示某一确定的时间段各相关数据的数值。而面板数据则是时间序列中不同时间段的截面数据的集合。

例如,表 1.2 所列的数据是时序数据,表示从 1978 年到 2004 年,中国能源生产总量的变化。

表 1.2 中国能源 1978—2004 年生产总量(单位:亿吨标煤)^①

年份	1978	1980	1985	1989	1990	1991	1992	1993	1994	1995
能源生产总量	6.2	6.3	8.5	10.1	10.3	10.4	10.7	11.1	11.8	12.9
年份	1996	1997	1998	1999	2000	2001	2002	2003	2004	
能源生产总量	13.2	13.2	12.4	10.9	10.6	12.0	13.8	15.9	18.4	

^① 数据来源:中国统计年鉴 2005,国家统计局编,中国统计出版社。

表 1.3 所列的数据是截面数据,表示 2004 年这一时间段中国能源生产的分类结构。

表 1.3 中国能源 2004 年分类生产量(单位:亿吨标煤)^①

能源分类	原煤	原油	天然气	水电
生产量	13.9	2.4	0.5	1.4

表 1.4 中的数据是面板数据,它表示 1978—2004 年原煤、原油、天然气和水电四类能源生产量的比例。其中每一列都是一个时间序列数据,每一行都是一个截面数据。由此可见,面板数据是按时间序列展开的截面数据集合。

表 1.4 中国能源 1978—2004 年分类生产量比例(单位:%)^②

年份	原煤	原油	天然气	水电
1978	70.3	23.7	2.9	3.1
1980	69.4	23.8	3.0	3.8
1985	72.8	20.9	2.0	4.3
1989	74.1	19.3	2.0	4.6
1990	74.2	19.0	2.0	4.8
1991	74.1	19.2	2.0	4.7
1992	74.3	18.9	2.0	4.8
1993	74.0	18.7	2.0	5.3
1994	74.6	17.6	1.9	5.9
1995	75.3	16.6	1.9	6.2
1996	75.2	17.0	2.0	5.8
1997	74.1	17.3	2.1	6.5
1998	71.9	18.5	2.5	7.1
1999	68.3	21.0	3.1	7.6
2000	66.6	21.8	3.4	8.2
2001	68.6	19.4	3.3	8.7
2002	71.2	17.2	3.1	8.4
2003	74.5	15.1	2.9	7.5
2004	75.6	13.5	3.0	7.9

4. 定类数据、定序数据、定距数据和定比数据

根据数据之间的关系,可以分为定类数据、定序数据、定距数据和定比数据。

(1) 定类数据:数据分为不同的类,不同类的数据具有不同的单位,对不同类的数据之间求和、计算平均值都是没有意义的。

在表 1.5 中,户总数(万户)和总人口数(万人)就是不同类的数据,计算某一个县的户总数和总人口数之和,或者计算户总数和总人口数的平均值,都是没有意义的。但总人口数(万人)和男(万人)、女(万人)就是同类数据。同类数据可以相加,同一个县的男人数和女人数之和就是总人口数,男人人口数和女人人口数之比就是该县人口的性别比。

^① 数据来源:中国统计年鉴 2005,国家统计局编,中国统计出版社。

^② 同上。

表 1.5 杭州市 2005 年人口统计表^①

	户总数(万户)	总人口数(万人)	男(万人)	女(万人)
全市	204.52	651.68	332.62	319.06
市区	120.56	401.59	204.43	197.16
桐庐县	14.32	39.44	20.10	19.34
淳安县	14.65	45.16	23.35	21.81
建德市	17.21	50.82	26.37	24.45
富阳市	20.20	62.78	32.05	30.73
临安市	17.59	51.89	26.31	25.57

在某些情况下,不同类别的数据之间可以计算比例,例如,每个县的总人口数与户总数之比,就是该县的户均人口数。

(2) 定序数据:数据根据一定的准则,排出一个数据序列。每一个数据都可以确定它在这个序列中的位置,如表 1.6 所示。

表 1.6 重庆地区主要城市环境评价指标^②

城市	重庆	德阳	绵阳	达县	雅安	西昌
环境评价价值(无量纲)	0.9883	0.1732	0.3172	0.7885	0.3713	0.6037
评价等级	三级	一级	二级	三级	二级	二级

在表 1.6 中,6 个城市环境指标和等级是一组定序数据,这组数据表明它们环境水平的排列次序,但这些评价数据之间的差距和比例是没有意义的。

(3) 定距数据:如果一组数值数据,可以比较它们之间的差距,但不能比较它们之间的比例,那么这组数据称为定距数据,如表 1.7 所示。

表 1.7 2005 年 12 个城市平均气温(°C)^③

序号	1	2	3	4	5	6	7	8	9	10	11	12
城市	北京	天津	石家庄	太原	呼和浩特	沈阳	大连	长春	哈尔滨	上海	南京	杭州
年平均气温	13.5	13.2	14.3	10.9	8.0	9.6	12.2	7.1	5.8	17.5	16.9	17.8

在表 1.7 中,各城市的平均气温之间的差距是有意义的,例如,上海的年平均气温比北京高 4°C。但气温的比值没有意义,因为 0°C 并不是气温的原点。

(4) 定比数据:对于一组数值数据,如果数值 0 是数据的原点,则数据之间可以计算比例。这样的数据称为定比数据,如表 1.8 所示。

表 1.8 2004 年 12 个城市房地产价格指数(2003 年价格为 100)^④

序号	1	2	3	4	5	6	7	8	9	10	11	12
城市	北京	天津	石家庄	太原	呼和浩特	沈阳	大连	长春	哈尔滨	上海	南京	杭州
价格指数	103.4	101.2	100.1	99.6	101.5	99.0	98.2	100.9	100.6	105.5	105	107.6

① 数据来源:杭州市统计局,《杭州统计年鉴》(2005)。

② 数据来源:白润才、殷伯良、孙庆宏,《BP 神经网络模型在城市环境质量评价中的应用》,载《辽宁工程技术大学学报》(自然科学版)2001 年第 3 期。

③ 数据来源:国家统计局,《中国统计年鉴》(2005)。

④ 同上。