

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

DATA, MODELS, AND DECISIONS

蒋绍忠 编著






北京大学出版社
PEKING UNIVERSITY PRESS

- ◎ 内容全面，包括概率和统计的基本概念、数据处理和分析的基本方法，以及回归、预测、风险分析、模拟和决策优化等模型方法。
- ◎ 基于Excel实现所有的模型和方法。本书提供大量的Excel操作截屏图，并详细加以标注，以方便读者学习和应用。
- ◎ 配有丰富的教辅资源。

本书适合MBA和工商管理学科研究生教学使用，也可作为企事业单位管理决策人员、数据分析人员自学和培训用书。

本书配套教辅资源

-  **PPT课件**，长达900多页的超大容量PPT，图文并茂，拿来即用，免费申请。
-  **学习光盘**，提供书中所有例题和习题的Excel文件，随书附赠。
-  **教师光盘**，长达数十个小时的课堂视频AVI文档以及所有习题解答的Excel文档。

上架建议：MBA·工商管理

ISBN 978-7-301-13805-2



9 787301 138052 >

定价：52.00元（附光盘）

TP391.13
J589

21世纪 **MBA** 规划教材

数据、模型与决策： 基于Excel的建模和商务应用

DATA, MODELS, AND DECISIONS

蒋绍忠 编著



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. —北京:北京大学出版社,2010.1

(21 世纪 MBA 规划教材)

ISBN 978 - 7 - 301 - 13805 - 2

I. 数… II. 蒋… III. 电子表格系统, Excel - 研究生 - 教材 IV. TP391.13

中国版本图书馆 CIP 数据核字(2008)第 066478 号

书 名: 数据、模型与决策——基于 Excel 的建模和商务应用

著作责任者: 蒋绍忠 编著

责任编辑: 张静波

标准书号: ISBN 978 - 7 - 301 - 13805 - 2/F · 1924

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

电子邮箱: em@pup.pku.edu.cn

印 刷 者: 三河市欣欣印刷有限公司

经 销 者: 新华书店

787 毫米 × 1092 毫米 16 开本 28 印张 705 千字

2010 年 1 月第 1 版 2010 年 1 月第 1 次印刷

印 数: 0001—4000 册

定 价: 52.00 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 电子邮箱:fd@pup.pku.edu.cn

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	(7)
1.3 数据透视表和模拟运算表	(11)
1.4 数据展示图	(20)
1.5 数据的分布属性	(29)
习题	(45)
第 2 章 概率论和统计学基础	(47)
2.1 随机事件、随机变量和频数	(49)
2.2 离散型随机变量及概率分布	(53)
2.3 连续型随机变量及概率分布	(60)
2.4 联合分布、边际分布和条件概率	(72)
2.5 统计学概述	(77)
习题	(78)
第 3 章 抽样和估计	(81)
3.1 统计抽样	(83)
3.2 样本数据的统计分析	(87)
3.3 点估计	(89)
3.4 区间估计和置信区间	(91)
3.5 置信水平、置信区间宽度和样本数	(101)
习题	(102)
第 4 章 假设检验	(105)
4.1 假设检验的基本原理	(107)
4.2 单样本假设检验	(115)
4.3 两样本假设检验	(120)
4.4 多样本假设检验方差分析	(126)
习题	(138)
第 5 章 回归分析	(141)
5.1 简单线性回归	(143)
5.2 一元线性回归	(147)

5.3	多元线性回归	(158)
5.4	非线性回归	(164)
5.5	具有属性变量的回归	(174)
5.6	正确运用回归分析方法	(177)
	习题	(177)
第 6 章	统计预测	(179)
6.1	预测概述	(181)
6.2	用回归方程预测	(182)
6.3	时间序列预测	(195)
6.4	预测工具 CB Predictor	(215)
	习题	(221)
第 7 章	风险决策和蒙特卡罗模拟	(223)
7.1	决策概述	(225)
7.2	风险决策	(226)
7.3	风险决策的效用理论	(234)
7.4	蒙特卡罗模拟	(239)
7.5	投资决策的蒙特卡罗模拟	(247)
	习题	(251)
第 8 章	风险分析工具 Crystal Ball	(253)
8.1	Crystal Ball 基础	(255)
8.2	Crystal Ball 在企业经营风险分析中的应用	(288)
8.3	Crystal Ball 在项目管理风险分析中的应用	(297)
	习题	(315)
第 9 章	管理系统模拟	(319)
9.1	模型和模拟	(321)
9.2	库存系统	(323)
9.3	库存系统模拟	(328)
9.4	排队系统的模拟	(335)
9.5	单服务台、单队列排队系统模拟	(346)
	习题	(351)
第 10 章	管理系统优化	(353)
10.1	线性规划优化数学模型	(355)
10.2	线性规划问题的基本概念	(362)
10.3	用规划求解工具求解线性规划问题	(365)
10.4	线性规划问题求解结果的分析	(377)
	习题	(385)

第 11 章 多目标决策	(387)
11.1 多目标决策概述	(389)
11.2 层次分析法	(391)
11.3 住宅选择的层次分析模型	(400)
11.4 层次分析法的 Excel 模型	(407)
习题	(411)
附录	(413)
附录 1 标准正态分布表	(415)
附录 2 累积标准正态分布表	(416)
附录 3 t 检验值	(418)
附录 4 χ^2 检验值	(421)
附录 5-1 F 检验值 ($\alpha = 0.05$)	(422)
附录 5-2 F 检验值 ($\alpha = 0.025$)	(424)
附录 5-3 F 检验值 ($\alpha = 0.005$)	(426)
附录 6 Excel 函数一览表	(428)

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。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℃”、“3 018.18 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℃”、“3 018.18 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. 时间序列数据和截面数据

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)和截面数据。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。例如,表 1.2 所列的数据是时序数据,表示从 1978 年到 2004 年,城乡家庭恩格尔系数的变化。

表 1.2 城乡居民家庭恩格尔系数(%)^①

年份	1978	1980	1985	1989	1990	1991	1992	1993	1994	1995
农村居民家庭	67.7	61.8	57.8	54.8	58.8	57.6	57.6	58.1	58.9	58.6
城镇居民家庭	57.5	56.9	53.3	54.5	54.2	53.8	53.0	50.3	50.0	50.1
年份	1996	1997	1998	1999	2000	2001	2002	2003	2004	
农村居民家庭	56.3	55.1	53.4	52.6	49.1	47.7	46.2	45.6	47.2	
城镇居民家庭	48.8	46.6	44.7	42.1	39.4	38.2	37.7	37.1	37.7	

表 1.3 所列数据是截面数据,表示 2005 年这一时间段内 5 个城市空气质量的主要指标。

表 1.3 城市空气质量指标(2005 年)^②

序号	城市	主要成分(毫克/立方米)			空气质量(天)	
		可吸入颗粒物	二氧化硫	二氧化氮	二级以上的天数	二级以上天数占全年比例
1	北京	0.141	0.050	0.066	234	64.11
2	天津	0.106	0.076	0.047	298	81.64
3	哈尔滨	0.104	0.042	0.056	301	82.47
4	上海	0.088	0.061	0.061	322	88.22
5	广州	0.088	0.053	0.068	332	90.96

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 1.4 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,男人口数和女人口数之比就是该县人口的性别比。

表 1.4 杭州市 2005 年人口统计表^③

	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

① 恩格尔系数是家庭食品支出占家庭总支出的比例。数据来源:国家统计局,《中国统计年鉴》(2005)。

② 数据来源:国家统计局,《中国统计年鉴》(2005)。

③ 数据来源:杭州市统计局,《杭州统计年鉴》(2005)。

这个序列中的位置,如表 1.5 所示。

表 1.5 重庆地区主要城市环境评价指标^①

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 1.5 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 1.6 所示。

表 1.6 2005 年 12 个城市平均气温(°C)^②

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 1.6 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 1.7 所示。

表 1.7 2004 年 12 个城市房地产价格指数(2003 年价格为 100)^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
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

很明显,表 1.7 中各城市的房地产价格指数是可以计算比例的,例如,2004 年杭州的房地产价格指数是沈阳的 1.087 倍。这组定比数据同时也是定序数据,但这些数据之间的差没有意义,因为它们的基数各不相同,因此不是定距数据。

根据以上的讨论,数据分类可以用图 1.1 表示。

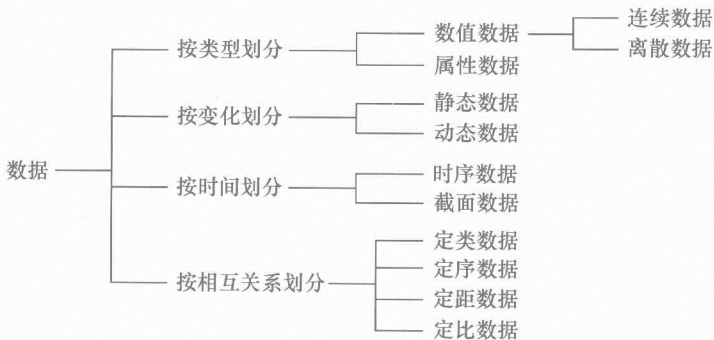


图 1.1 数据的分类

① 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。

② 数据来源:国家统计局,《中国统计年鉴》(2005)。

③ 同上。

1.1.3 数据收集和处理的道德准则和行为规范

数据的客观性和可重复性是数据收集和处理的的基本原则,商业数据的收集和处理也必须遵循这些原则。如果违反这些基本原则,获得的数据会造成对实际情况的判断错误,导致决策失误。商业数据收集和处理应该遵循以下行为规范:

(1) 应尽可能收集原始数据或称第一手收据,即收集那些经济活动和企业生产经营中产生的、未经处理和变动的数据记录。在数据分析报告中应说明原始数据来源、时段。

(2) 如果需要引用别人提供或处理过的数据,首先应该辨别是否侵犯数据提供者的知识产权。在合法引用的前提下,应该详细注明所引用数据的来源,既表示对数据提供者工作的尊重,也便于第三者需要时进一步考证。窃取别人提供的数据,作为自己的成果,是违背学术道德规范的行为。

(3) 原始数据中如果有异常数据,应尽可能分析产生异常的原因,并在数据分析报告中加以说明。如果异常数据是孤立或偶然的原因产生的,数据处理中有必要删除这些异常数据,则需要特别加以说明。为了达到事先设定的主观目的,随意增删、篡改客观数据,是违背学术道德规范的行为。

(4) 公布数据处理结果时,应该明确说明数据处理的工具、方法以及数据处理的流程,以便别的数据研究者 and 使用者确信,有关数据经过这样的处理,确实可以得到相应的结论。

(5) 通过问卷调查得到的数据,必须说明问卷调查内容、问卷调查表的格式、确定问卷调查对象的原则和方式、问卷发放和回收的渠道和方式、问卷发放数量、回收数量、有效问卷的数量、问卷结果统计方法和统计结果等。

总而言之,商业数据收集和处理与对待科学数据一样,是一项重要的基础工作,必须坚持实事求是的科学精神和严谨务实的科学态度。只有数据收集和处理工作做好了,数据的研究和分析才会有意义。

1.2 Excel 基础

Excel 电子表格是 Microsoft Office 的一个组件。由于 Microsoft 公司 Windows 操作系统具有压倒性的市场占有率,Excel 和其他 Office 组件一样,以其完善的功能、友好的界面和稳定的性能成为目前应用最广泛的电子表格工具。

Microsoft Office 在发展过程中,形成了 Office 97、Office 2000、Office XP、Office 2003、Office 2007 等版本。这些版本虽然不断改进,但主要功能和界面差别不大。Office 2007 的功能和界面变化较大,由于该版本目前还不太普及,本书以 Office 2003 为基础介绍 Excel 的应用。

阅读本书,需要掌握 Excel 的基本概念和基本操作。接下来,我们将介绍这些基本概念和基本操作。

1.2.1 工作簿和工作表

默认的 Excel 工作簿由“Sheet 1”、“Sheet 2”和“Sheet 3”三张工作表组成,如图 1.2 所示。

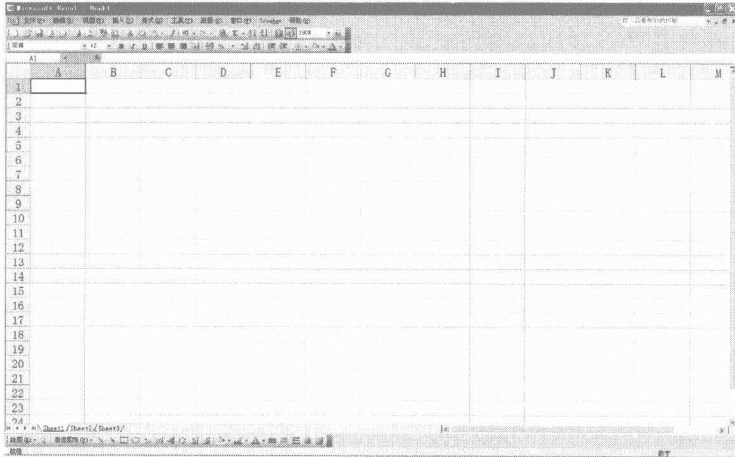


图 1.2 Excel 工作簿和工作表

右键单击工作表名,弹出图 1.3 所示菜单。

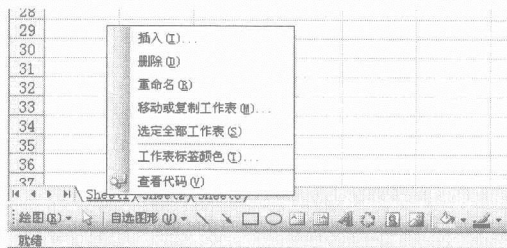


图 1.3 工作表的插入、删除和重命名

单击这些菜单项,可以插入、删除或重命名工作表。Excel 最多可以包含 255 张工作表。

1.2.2 工作表的行列和单元格

每个 Excel 工作表行的标号为 1、2、3……最多可以输入 65 536 行。列的标号为 A—Z, AA—AZ, BA—BZ, …, HA—HZ, IA—IV, 最多可以输入 256 列。

Excel 的单元格是 Excel 表的基本元素。单元格的格式可以用“格式”|“单元格”命令来定义。单元格的格式类型如图 1.4 所示。