



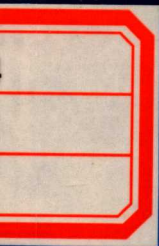
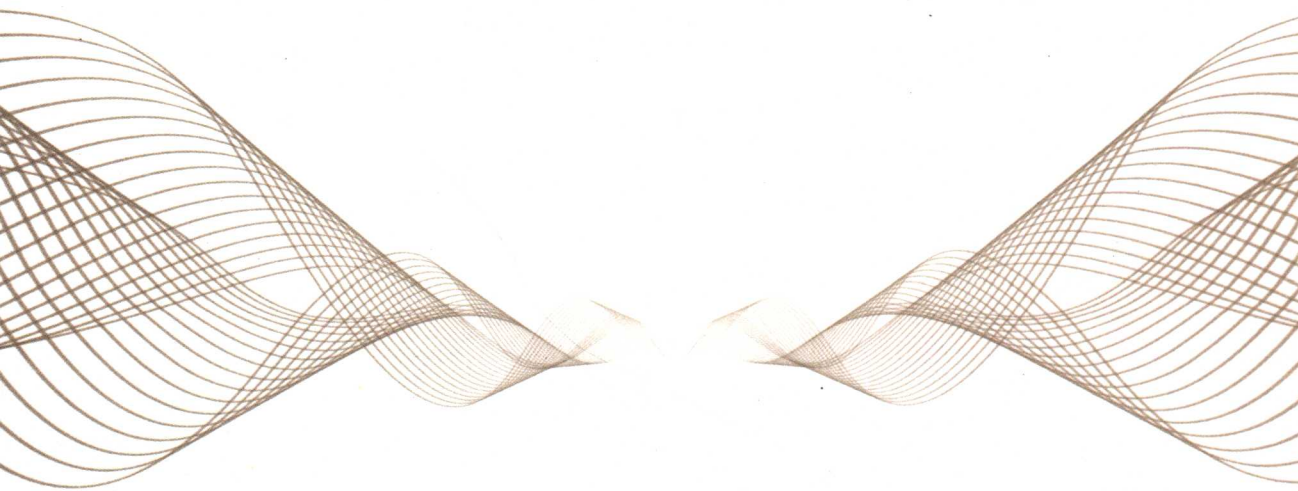
西安电子科技大学研究生精品教材

*Data, Models and Decisions*

# 数据、模型与决策

主 编 姜晓兵

副主编 谢永平 王 涛



西安电子科技大学出版社  
<http://www.xduph.com>

内容简介

西安电子科技大学研究生精品教材

# 数据、模型与决策

主 编 姜晓兵  
副主编 谢永平 王 涛

西安电子科技大学出版社

## 内 容 简 介

本书采用经济、管理人员易接受的方式,以数据—模型—决策为主线,将相关知识和方法有机地联系起来。全书共十章,主要内容包括数据、模型与决策基础,数据收集,数据整理,数据分布特征描述,数据推理基础:概率分布与抽样分布,数据推理方法:参数估计与假设检验,多总体均值比较模型:方差分析,因果关系研究模型:相关与回归分析,决策分析。本书注重与 Excel、SPSS 软件相结合,给出了软件分析实现的操作步骤和结果。为方便学习,每章内容配有引例、学习目标、本章小结、思考练习以及丰富的案例等。

本书可作为经济、管理类专业研究生教材,尤其适合作为工商管理硕士(MBA)、公共管理硕士(MPA)、工程硕士(ME)等专业培养中定量分析课程的教材,还可作为经济、管理类专业本科生的教材或教学参考书,对广大经济、管理领域的实际工作者也极具参考价值。

## 图书在版编目(CIP)数据

数据、模型与决策/姜晓兵主编. —西安:西安电子科技大学出版社,2017.1  
西安电子科技大学研究生精品教材  
ISBN 978-7-5606-4348-9

I. ① 数… II. ① 姜… III. ① 决策模型  
IV. ① C934

中国版本图书馆 CIP 数据核字(2017)第 001777 号

策划编辑 戚文艳

责任编辑 杨 璠

出版发行 西安电子科技大学出版社(西安市太白南路 2 号)

电 话 (029)88242885 88201467 邮 编 710071

网 址 www.xduph.com 电子邮箱 xdupfxb001@163.com

经 销 新华书店

印刷单位 陕西大江印务有限公司

版 次 2017 年 1 月第 1 版 2017 年 1 月第 1 次印刷

开 本 787 毫米×1092 毫米 1/16 印张 15.5

字 数 364 千字

印 数 1~2000 册

定 价 31.00 元

ISBN 978-7-5606-4348-9/TC

**XDUP 4640001-1**

\*\*\* 如有印装问题可调换 \*\*\*

本社图书封面为激光防伪覆膜,谨防盗版。

# 前 言

数据无孔不入，大数据时代，谁掌握了数据，谁就能把握成功。“一切皆可量化”，数据分析专家道格拉斯这个大胆的断言指出了解决诸多经济、管理活动问题的关键所在。

为传播数据、模型、决策的定量分析思维方式，本书在对数据、模型与决策理论知识、方法进行严格逻辑推理的基础上，加入了对数学表达式的解释，使其背后的现实经济、管理意义能够为人们所理解、熟悉，应用于自己的生活、工作中。

与同类教材相比较，本书的主要特征表现为：

(1) 注重突出数据分析方法应用技能和技巧的训练。本书为方便学生学习与理解，对所涉及的方法产生的背景、数学上的提法、适用的范围和对象、具体解题的步骤进行了详细的介绍，并结合相应的事例进行说明。直接面向问题，体现大数据时代“知道‘是什么’比知道‘为什么’更重要”、“相关关系比因果关系更重要”的思维、实践新变化。

(2) 注重突出数据分析方法操作的可行性，着重提升学生解决实际问题的能力。为此，本书依据实际问题，讲授了如何使用 Excel、SPSS 两种软件进行具体分析。借助计算机软件的帮助，能够快速处理大量的、复杂的数据资料，也使模型的求解变得更容易。

(3) 除使用大量应用具体方法分析的问题导向型案例外，还编入了使用系统方法才能解决的综合案例。通过这种安排，一方面可以更具具体地了解相关方法的原理和应用过程，另一方面也为将数据分析方法运用于专业问题研究提供了基本的参考。

为方便教师教学和学生自学，增加阅读的可读性与趣味性，每章内容均配有引例、学习目标、本章小结和思考练习。同时，在相应的章节适时地配备了综合案例，并给出了案例使用说明，以方便读者熟悉、理解和应用所讲的理论、知识与方法。根据授课对象或课时安排的不同，可合理选择适当内容讲授和学习。

编者从事数据分析教学与科研工作已十载，有道是“十年磨一剑”。在此期间有所得，有所失，收获了学生、同行的认可和自己专业知识的增进。近日遇到一学生提起当年在课堂上听到的我所谈论过的一些言语，发现自己都不曾记得了，愕然于莫言那句“用嘴说出的话随风而散，用笔写出的话永不磨灭”。在哈佛大学箴言：“不出版，就死亡”(Publish or Perish)的激励下，提笔把多年教学和科研上的些许收获编入此书中，供大家分享，但难免仍有“野人献曝”之态，只是想说明仍在追求“不滞于物，草木竹石均可为剑”的路上，亦想判断“在人的一生中，总有一个记忆力方兴未艾，而理解力运若转轴的最好时期”，自己是否错过。

本书的编写获得了西安电子科技大学研究生精品教材建设项目基金的资助，得到了西安电子科技大学经济与管理学院、研究生院、高级管理教育中心等有关部门领导和同事的大力支持。编写过程中，参阅、借鉴了大量国内外相关文献资料和同类教材，主要的文献资料和

教材已列于书后，在此向所有相关的作者表示深深的感谢！西安电子科技大学出版社戚文艳等诸位编辑为本书的出版付出了大量的汗水和辛勤的劳动，在此一并向他们致以诚挚的谢意。

本书是编者多年从事数据、模型与决策课程教学的成果。然而，鉴于编者才疏学浅，书中不妥之处在所难免，恳请各位专家、学者、教师、学生以及广大读者将宝贵的批评、意见和建议及时反馈给我们，对此我们表示衷心的感谢！编者的电子邮箱为 [xdjiangxiaobing@qq.com](mailto:xdjiangxiaobing@qq.com)，恭请各位不吝赐教。

编者

2016年7月

# 目 录

<b>第 1 章 数据、模型与决策基础</b> .....	1	<b>第 3 章 数据整理</b> .....	39
引例 数的起源 .....	1	引例 二战期间轰炸机战损率降低策略 .....	39
学习目标 .....	1	学习目标 .....	39
1.1 随处可见的数据 .....	2	3.1 分组与频数分布 .....	39
1.1.1 数字化运动:慢跑 .....	2	3.1.1 分组 .....	40
1.1.2 数字化阅读:红楼梦作者考证 .....	2	3.1.2 频数分布 .....	41
1.1.3 数字化营销:啤酒与尿布 .....	3	3.2 品质数据整理与图示 .....	41
1.1.4 数字化管理:六西格玛管理 .....	3	3.2.1 品质数据整理 .....	41
1.2 数据、模型与决策的内涵 .....	4	3.2.2 品质数据图示 .....	44
1.2.1 数据 .....	4	3.3 数值数据整理与图示 .....	46
1.2.2 模型 .....	7	3.3.1 数值数据整理 .....	46
1.2.3 决策 .....	9	3.3.2 数值数据图示 .....	51
1.2.4 数据、模型与决策的关系 .....	11	3.4 顺序数据、多指标数据与多总体数据 整理与图示 .....	55
1.3 常用的数据分析软件 .....	11	3.4.1 顺序数据整理与图示 .....	55
1.3.1 Excel 软件 .....	11	3.4.2 多指标数据整理与图示 .....	57
1.3.2 SPSS 软件 .....	13	3.4.3 多总体数据整理与图示 .....	58
1.3.3 SAS 软件 .....	15	3.5 统计表与统计图合理使用 .....	59
1.3.4 Stata 软件 .....	15	3.5.1 统计表的结构与设计 .....	59
本章小结 .....	15	3.5.2 统计图的绘制 .....	60
思考练习 .....	16	本章小结 .....	60
<b>第 2 章 数据收集</b> .....	17	思考练习 .....	61
引例 麦当劳选址的商圈调查 .....	17	<b>第 4 章 数据分布特征描述</b> .....	64
学习目标 .....	18	引例 平均人 .....	64
2.1 数据来源 .....	18	学习目标 .....	64
2.1.1 直接数据 .....	18	4.1 数据分布集中趋势 .....	65
2.1.2 间接数据 .....	27	4.1.1 数值平均数 .....	65
2.2 问卷设计 .....	28	4.1.2 位置平均数 .....	70
2.2.1 问卷基本结构 .....	28	4.1.3 综合平均数:截尾均值 .....	73
2.2.2 提问项目设计 .....	31	4.1.4 算术平均数、众数与中位数间的关系 .....	74
2.2.3 回答项目设计 .....	32	4.2 数据分布离中趋势 .....	75
2.2.4 问题顺序设计 .....	35	4.2.1 异众比率 .....	75
2.3 数据误差与质量 .....	36	4.2.2 全距与四分位差 .....	75
2.3.1 数据误差 .....	36	4.2.3 平均差 .....	76
2.3.2 数据质量 .....	36	4.2.4 方差与标准差 .....	77
本章小结 .....	37		
思考练习 .....	37		

4.2.5 离散系数 .....	79	6.2 总体均值区间估计 .....	121
4.2.6 算术平均数与标准差的结合使用 .....	79	6.2.1 单总体均值的区间估计 .....	121
4.3 数据分布形态度量 .....	81	6.2.2 双总体均值差的区间估计 .....	125
4.3.1 偏度 .....	81	6.3 总体方差区间估计 .....	130
4.3.2 峰度 .....	82	6.3.1 单总体方差的区间估计 .....	130
4.3.3 数据分布特征的软件操作 .....	83	6.3.2 双总体方差比的区间估计 .....	131
本章小结 .....	84	本章小结 .....	132
思考练习 .....	84	思考练习 .....	133
综合案例 职场上的年龄歧视 .....	86		
<b>第5章 数据推理基础：概率分布与抽样分布</b> .....	90	<b>第7章 数据推理方法：假设检验</b> .....	135
引例 个体与个性的产生 .....	90	引例 女士品茶 .....	135
学习目标 .....	91	学习目标 .....	135
5.1 概率与概率分布 .....	91	7.1 假设检验的原理 .....	136
5.1.1 概率的获得 .....	91	7.1.1 假设检验的思路 .....	136
5.1.2 概率的运算 .....	92	7.1.2 假设检验的步骤 .....	136
5.1.3 概率分布 .....	93	7.2 总体均值假设检验 .....	140
5.2 典型离散型随机变量的概率分布 .....	96	7.2.1 单总体均值的假设检验 .....	140
5.2.1 伯努利分布 .....	96	7.2.2 双总体均值差的假设检验 .....	144
5.2.2 二项分布 .....	96	7.3 总体方差假设检验 .....	150
5.2.3 泊松分布 .....	98	7.3.1 单总体方差的假设检验—— $\chi^2$ 检验 .....	150
5.2.4 超几何分布 .....	98	7.3.2 双总体方差比的假设检验—— $F$ 检验 .....	151
5.3 典型连续型随机变量的概率分布 .....	99	本章小结 .....	152
5.3.1 均匀分布 .....	99	思考练习 .....	153
5.3.2 指数分布 .....	100		
5.3.3 正态分布 .....	101	<b>第8章 多总体均值比较模型：方差分析</b> .....	155
5.3.4 推断分布 .....	104	引例 耐克公司的高尔夫球广告 .....	155
5.4 抽样分布 .....	107	学习目标 .....	156
5.4.1 单样本均值的抽样分布 .....	108	8.1 方差分析原理 .....	156
5.4.2 双样本均值差的抽样分布 .....	108	8.1.1 方差分析概念 .....	156
5.4.3 单样本方差的抽样分布 .....	109	8.1.2 方差分析思路 .....	158
5.4.4 双样本方差比的抽样分布 .....	109	8.1.3 方差分析假设 .....	159
本章小结 .....	110	8.2 单因素方差分析 .....	159
思考练习 .....	111	8.2.1 单因素方差分析问题描述 .....	159
综合案例 意外伤害险的赔付 .....	112	8.2.2 单因素方差分析假设检验 .....	161
		8.3 双因素方差分析 .....	165
<b>第6章 数据推理方法：参数估计</b> .....	114	8.3.1 无交互作用的双因素方差分析 .....	165
引例 二战期间德军坦克数量估计 .....	114	8.3.2 有交互作用的双因素方差分析 .....	170
学习目标 .....	115	本章小结 .....	176
6.1 参数估计原理 .....	115	思考练习 .....	176
6.1.1 点估计 .....	115	综合案例 不同地区公司盈利性差异 .....	179
6.1.2 区间估计 .....	119		

<b>第9章 因果关系研究模型：相关与回归分析</b> .....	181	学习目标 .....	212
引例 高尔顿与回归分析的起源 .....	181	10.1 决策基础 .....	213
学习目标 .....	182	10.1.1 决策原则 .....	213
9.1 相关分析与回归分析 .....	182	10.1.2 决策类型 .....	214
9.1.1 变量间的关系 .....	182	10.1.3 决策分析 .....	216
9.1.2 相关关系的描述 .....	183	10.2 非确定型决策 .....	217
9.1.3 相关关系的测度 .....	186	10.2.1 乐观准则 .....	218
9.1.4 回归分析 .....	189	10.2.2 悲观准则 .....	218
9.2 一元线性回归分析 .....	190	10.2.3 折中准则 .....	219
9.2.1 一元线性回归模型的描述 .....	190	10.2.4 后悔值准则 .....	220
9.2.2 一元线性回归方程的构建 .....	191	10.2.5 等可能准则 .....	221
9.2.3 一元线性回归方程的检验 .....	194	10.3 风险型决策 .....	222
9.2.4 一元线性回归方程的预测 .....	199	10.3.1 期望收益值准则 .....	222
9.3 多元线性回归分析 .....	202	10.3.2 期望效用值准则 .....	223
9.3.1 多元线性回归模型的描述 .....	202	10.3.3 决策树 .....	226
9.3.2 多元线性回归方程的构建 .....	203	本章小结 .....	228
9.3.3 多元线性回归方程的检验 .....	203	思考练习 .....	228
9.3.4 多元线性回归方程的应用 .....	205	附录1 标准正态分布函数值表 .....	232
本章小结 .....	206	附录2 $\chi^2$ 分布临界值表 .....	233
思考练习 .....	207	附录3 $t$ 分布临界值表 .....	234
综合案例 销售额影响因素 .....	209	附录4 $F$ 分布临界值表( $\alpha=0.05$ ) .....	235
<b>第10章 决策分析</b> .....	212	参考文献	238
引例 狗熊来了 .....	212		



## 第1章 数据、模型与决策基础

### 引例

#### 数的起源

数,是我们在生活中常常接触的。数无处不在,买东西,商品的价钱要用到数;看电视,电视频道的标号要用到数;写封信,填邮政编码还要用到数。可见数在我们生活中是非常重要的。

原始时代的人类,为了维持生活必须每天外出狩猎和采集果实。有时他们满载而归,有时却一无所获;带回的食物有时有富余,有时却不足果腹。生活中这种数与量上的变化使人类逐渐产生了数的意识。在那个时候,人们在生产活动中注意到一只羊与许多羊、一头狼与整群狼在数量上的差异,开始了解有与无、多与少的差别,进而知道了一和多的区别,然后又明确了从多到二、三等单个数目概念。随着社会的进一步发展,简单的计数就是必不可少的了,一个部落集体必须知道它有多少成员或有多少敌人,一个人也必须知道他的羊群里的羊是不是少了。这样,人类的祖先在与大自然的艰难搏斗中,在漫长的生活实践中,由于记事和分配生活用品等方面的需要,逐渐产生了数的概念。数的概念的形成可能与火的使用一样古老,大约是在30万年以前,它对于人类文明的意义也不亚于火的使用。数的产生,标志着人类的思维逐步由事件的直观思维走向形式或抽象思维。

最早人们利用自己的十个指头来记数,当指头不敷应用时,人们开始采用“石头记数”、“结绳记数”和“刻痕记数”。在经历了数万年的发展后,直到距今大约五千多年前,才出现了书写记数以及相应的记数系统。早期记数系统有:公元前3400年左右的古埃及象形数字;公元前2400年左右的巴比伦楔形数字;公元前1600年左右的中国甲骨文数字;公元前500年左右的希腊阿提卡数字;公元前500年左右的中国筹算数字;公元前300年左右的印度婆罗门数字以及年代不详的玛雅数字。这些记数系统采用不同的进制,其中除巴比伦楔形数字采用六十进制、玛雅数字采用二十进制外,其他均采用十进制,从古到今数的运算很大程度上是依靠了十进位制的伟大发明。当今几乎主宰了人们工作和生活一切领域的计算机则是得益于数字的二进位制才形成了它的千变万化。



#### 学习目标

- 了解数据、模型与决策分析方法的应用;
- 掌握数据、模型与决策的基本概念;
- 掌握数据、模型与决策三者间的相互关系;
- 熟悉模型使用中管理人员承担的工作;
- 了解常用的数据分析软件。

## 1.1 随处可见的数据

当今，数据在我们的生活、工作中随处可见。这正印证了一百多年前英国作家赫伯特·乔治·威尔斯(Herbert George Wells)给出的预言，数据分析、推理方法总有一天会像读与写一样成为所有人必须具备的能力。

### 1.1.1 数字化运动：慢跑

人体的最佳活动时间应是 18:00 至 20:00 点。因为人体的各种活动都受到“生物钟”的控制，在一天 24 小时内，人体力的最高点和最低点都有一定的规律性，而绝大多数人体力发挥的最高点并不在清晨，而是在傍晚。傍晚或者夜幕降临之时出来跑步，白天工作上的困扰能通过夜跑获得精神上的放松，朦胧的夜色也能让压抑的心灵复归平静。如果你在路上看见一个左手腕戴着 GPS 腕表、右手腕绕着一条运动跟踪腕带、一副运动耳机连至运动裤兜里的智能手机的慢跑者——那就是我，一个“数字化”慢跑者的标准形象。

一名“Geek”大概是不能忍受一样没有任何科技含量的活动的——所以我的夜跑运动计划的第一项内容并不是规划慢跑路线，计算运动量，而是买齐所有最新最酷的运动辅助数码科技装备——没必要规划路线，GPS 腕表能够帮助我记录下我的跑步路线，它还能提供实时心律，以及“和昨天的自己赛跑”的功能；运动量也没必要计算，运动跟踪腕带能够跟踪记录我一天中消耗的所有卡路里，它还能记录我的睡眠状况；联网的电子秤把每日体重数据发到我的手机上——所有数据都会汇总到智能手机上的 App 中，运动情形、健康状况一目了然(见图 1.1)。

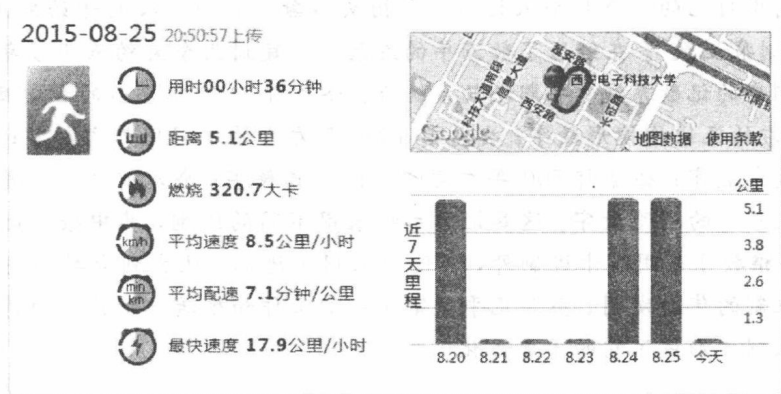


图 1.1 一名慢跑者一次运动的部分数据

### 1.1.2 数字化阅读：红楼梦作者考证

众所周知，《红楼梦》一书共 120 回，自从胡适作《红楼梦考证》以来，一般都认为前 80 回为曹雪芹所写，后 40 回为高鹗所续，然而长期以来这种看法一直都饱受争议。从 1985 年开始，复旦大学的李贤平教授带领他的学生进行了数据分析的考证工作，他们创造性的想法是将 120 回看成是 120 个样本，然后确定与情节无关的虚词出现的次数作为变量，巧妙运用数据分析方法，看看哪些回目出自同一人的手笔。一般认为，每个人使用一些词的习惯是特有

的。于是李教授用每个回目中47个虚词(之、其、或、亦……呀、吗、咧、罢……可、便、就等)出现的次数(频率),作为《红楼梦》各个回目的数字标志。之所以要抛开情节,是因为在一般情况下,同一情节大家描述的都差不多,但由于个人写作特点和习惯的不同,所用的虚词是不会一样的。利用多元分析中的聚类分析法进行聚类,果然将120回分成两类,即前80回为一类,后40回为一类,很形象地证实了《红楼梦》一书不是出自同一人的手笔。之后又进一步分析前80回是否为曹雪芹所写。这时又找了一本曹雪芹的其他著作,进行了类似计算,结果证实了用词手法完全相同,断定前80回出自曹雪芹一人手笔,是他根据《石头记》写成,中间插入《风月宝鉴》,还有一些其他成分;而后40回是否为高鹗写的呢?论证结果推翻了后40回是高鹗一个人所写,而是曹雪芹亲友将其草稿整理而成,宝黛故事为一人所写,贾府衰败情景当为另一人所写,等等。这个论证在红学界轰动很大,李教授他们用多元统计分析方法支持了红学界的观点,使红学界大为赞叹。

即使续写作者刻意模仿曹雪芹的写法,但是文风是不能模仿的,对虚词的使用更是难以做到一致,这就是标准(也就是统计量)所在。李教授的工作便是证明了前80回和后40回在虚词的使用上截然不同。而《石头记》与《风月宝鉴》的对比使用的则是因子分析的方法。每一回47个虚词出现不同次数,而一共有120回,这样就构成一个 $47 \times 120$ 的矩阵,李教授运用统计软件分析这个大型矩阵得到以上结果,可信度甚高,因为它是完全客观的方法。

### 1.1.3 数字化营销:啤酒与尿布

“啤酒与尿布”的故事发生于20世纪90年代的美国沃尔玛超市中,沃尔玛的超市管理人员分析销售数据时发现了一个令人难以理解的现象:在一些特定的情况下,“啤酒”与“尿布”两件看上去毫无关系的商品会经常出现在同一个购物篮中,这种独特的销售现象引起了管理人员的注意,经过后续调查发现,这种现象出现在年轻的父亲身上。

在美国有婴儿的家庭中,一般是母亲在家中照看婴儿,年轻的父亲前去超市购买尿布。父亲在购买尿布的同时,往往会顺便为自己购买啤酒,这样就会出现啤酒与尿布这两件看上去不相干的商品经常会出现在同一个购物篮的现象。如果这个年轻的父亲在卖场只能买到两件商品之一,则他很有可能会放弃购物而到另一家商店,直到可以一次同时买到啤酒与尿布为止。沃尔玛发现了这一独特的现象,开始在卖场尝试将啤酒与尿布摆放在相同的区域,让年轻的父亲可以同时找到这两件商品,并很快地完成购物。沃尔玛超市可以让这些客户一次购买两件商品而不是一件,从而获得了很好的商品销售收入,这就是“啤酒与尿布”故事的由来。

当然“啤酒与尿布”的故事必须具有技术方面的支持。1993年,美国学者艾格拉沃(Agrawal)提出通过分析购物篮中的商品集合,从而找出商品之间关联关系的关联算法,并根据商品之间的关系,分析客户的购买行为。艾格拉沃从数学及计算机算法角度提出了商品关联关系的计算方法——Apriori算法。沃尔玛从20世纪90年代尝试将Apriori算法引入到POS机数据分析中,并获得了成功,于是产生了“啤酒与尿布”的故事。

### 1.1.4 数字化管理:六西格玛管理

六西格玛( $6\sigma$ )概念作为品质管理概念,最早是由摩托罗拉公司的比尔·史密斯(Bill Smith)于1986年提出的,其目的是设计一个目标:在生产过程中降低产品及流程的缺陷次

数,防止产品变异,提升品质。而它真正流行并发展起来,源于通用电气公司的实践。20世纪90年代,通用电气公司在时任CEO的杰克·韦尔奇(Jack Welch)带领下,总结了全面质量管理的成功经验,提炼了其中流程管理技巧的精华和最行之有效的办法,使六西格玛管理发展成熟为一种提高企业业绩与竞争力的管理模式。杰克·韦尔奇曾谈到:“6 $\sigma$ 培训计画是GE下一个世纪领导层得以产生繁衍的园地,6 $\sigma$ 是我们曾经尝试过的最重要的管理培训方法,它胜过到哈佛工商学院就读,也胜过到克劳顿管理学院(克劳顿管理学院是GE高级管理人员培训中心,有人把它称为GE高级领导干部成长的摇篮,有‘美国企业界的哈佛’之誉)进修,它教会你一种完全与众不同的思维方式。”

西格玛即希腊字 $\sigma$ 的音译,是数据分析人员用于衡量工艺流程中的波动性而使用的代码。企业可以用西格玛的级别来衡量其在商业流程管理方面的表现。传统公司的一般品质要求已提升至3 $\sigma$ 。这就是说产品的合格率已达99.73%的水平,只有0.27%为次品,又或者解释为每生产1000件产品或提供1000次服务中只有2.7件或2.7次存有缺陷。很多人认为产品达至此水平已非常满意。可是,根据专家的研究结果证明,如果产品达到99.73%的合格率,以下事件便会继续在现实中发生:每年有20000次配错药事件;每年有15000名婴儿出生时会被抛落地上;每年平均有9小时没有水、电、暖气供应;每星期有500宗做错手术事件;每小时有2000封信邮寄错误。然而,采用6 $\sigma$ 标准,就产品生产或服务而言,则意味着在生产或服务过程中缺陷率不超过百万分之3.4,通俗地说,如果生产100万件产品或进行100万次服务,产品的不合格率或服务达不到要求的比率平均来说不超过3.4件或3.4次。这样的不合格率非常低,故可以忽略不计。

由此可以看出,随着人们对产品质量要求的不断提高和现代生产管理流程的日益复杂化,企业越来越需要6 $\sigma$ 这样的高端流程质量管理标准,以保持激烈的市场竞争中的优势地位。该管理方法在摩托罗拉、通用电气、戴尔、惠普、西门子、索尼、东芝、华硕等众多跨国企业的实践中证明是卓有成效的。为此,国内一些部门和机构在国内企业大力推行6 $\sigma$ 管理工作,引导企业开展6 $\sigma$ 管理。

由此判断,要列举出一个数据分析、推理方法没有用途的领域,是非常困难的。因为几乎所有的领域都会用到这种方法。作为管理人员,更应该掌握这种数据分析与推理的方法。在商业活动中,如果说以往凭经验甚至凭勇气还能多少获得一些利润,取得一些成就,那么,在经济高速发展、全球化日趋明显的今天,管理人员都深刻地意识到,竞争已经在更高的层面展开,光有笼统的、定性的、模糊的认识,没有精确的、定量的、深刻的分析将会方向不明、目标不清、行动不力,使自己处在极其不利的地位。这就是数据分析、推理方法越来越受到现代管理人员重视的原因。

## 1.2 数据、模型与决策的内涵

### 1.2.1 数据

数据,就其词义而言,是指进行各种统计、计算、科学研究和技术设计等所依据的数值。英语里data作为datum一词的复数形式,具有如下涵义:①论据,作为论据的事实;②材料,资料;③历史上的、计算或实验得到的数据。经济、管理领域分析研究的“数据”是作为集合名词出现的,是指进行决策所依据的客观事实基础。数据和信息是不可分离的,数据是

信息的表达,信息是数据的内涵,因此,才有了 data 与 information 互为近义词。数据本身没有意义,数据只有对实体行为产生影响时才成为信息,其反映的是一定社会现象或自然现象在特定的时间、空间条件下表现出的特征,故经济、管理领域分析研究的数据需要具备时间与空间要素,并非数学学科中纯粹的、抽象的数字。

数据的载体是客观现象,全部研究现象的集合称为总体。总体是根据研究目的而确定的、客观存在的,是在同一性质基础上由众多个别事物结合起来的整体。例如,研究全国服务业企业的利润情况时,全国所有的服务业企业形成了一个总体。成千上万不同的服务业企业可以结合在一起构成总体,这是因为每个服务业企业的经济职能是相同的,都是从事生产和销售服务产品的基本单位,在经营过程中都需要投入一定的成本以获取相应的收益。同质性是确定总体的基本标准,它是根据研究目的而确定的。研究目的不同,则所确定的总体也不同,其同质性的意义也随之变化。例如,研究全国服务业企业的营业收入时,全国所有的服务业企业构成了总体,凡是服务业企业都是同质的。如果研究的是中小型规模服务业企业的营业收入状况,那么,年营业收入在 2000 万元以下的服务业企业则构成了总体,这些企业是同质的,而年营业收入在 2000 万元以上的服务业企业就是非同质的了,不能归入此总体中。

总体单位(简称单位)是构成总体的各个事物,亦称个体。根据总体中包含的总体单位数目,总体可以分为有限总体和无限总体。总体所包含的个体数目是有限的,称为有限总体,如某一时刻一定区域内的人口、企业、汽车等。总体所包含的个体数目是无限的,称为无限总体,如一企业生产的某类产品,由于该产品可能会一直持续生产下去,总体包括已经生产、正在生产和将来生产的此类产品,数量难以具体确定,故属于无限总体。总体和总体单位的概念是相对而言的,随着研究目的改变而变化。同一个研究对象,在一种情况下为总体单位,但在另一种情况下有可能变成总体。例如,研究全国劳动密集型企业的员工结构,全国所有此类企业形成总体,每个劳动密集型企业为总体单位;但当研究局限于某一劳动密集型企业的员工结构时,则该企业就变成了总体,企业各个员工成了总体单位。

数据研究的目的是要确定总体的内在数量规律性。然而,当总体单位数量很多甚至无限时,不可能亦没有必要对构成总体的所有单位都进行调查。这时,采用一定的方式从总体中抽取一部分单位,作为总体的代表加以研究。这种由总体的部分单位组成的集合称为样本。样本和总体的关系主要体现在两方面。一方面,总体是人们所要研究的对象,而样本则是所要观察的对象,样本是总体的代表和缩影;另一方面,样本是用来推断总体的,对样本进行观察的目的是要对总体数量特征作出估计或判断,即通常所说的以样本推断总体。样本所包含的总体单位数目称为样本容量。

标志是总体各单位具有的共同特征的名称。从不同角度考察,每个总体单位可以有許多特征。例如,企业员工都具有性别、年龄、民族、受教育程度、工资、津贴等,这些都是员工的标志。标志在总体各单位中的具体表现即为数据。数据研究就是从收集各单位标志具体表现状况开始的,通过对标志的综合分析反映出总体的数量特征。

数据分析研究需要进行必要的计量,这就必须要弄清楚数据具备怎样的运算功能。依据美国社会学家、统计学家史蒂文斯(S. S. Stevens)对数据运算功能的研究,按照所具备的运算功能由低到高、由粗略到精确,将数据划分为四个类型:定类数据、定序数据、定距数据和定比数据,相应地,标志也分为四类:定类标志、定序标志、定距标志和定比标志(见表 1.1)。

表 1.1 标志与数据的类型

标志类型	数据类型	运算功能	特征	实例	
品质标志	定类标志	定类数据	计数	分类	性别、民族、品种
			计数	分类	等级、产业、优劣、
	定序标志	定序数据	排序	排序	奖励
数值标志	定距标志	定距数据	计数	分类	产品、质量、差异
			排序	排序	
	定比标志	定比数据	加减	基本的测量单位	产量、产值、商品 销售额、利润率、 增长率、人均数
			计数	分类	
			排序	排序	
		加减	基本的测量单位		
		乘除	绝对零点		

观察人的性别、企业所属的行业、学生所在的学院等，这里的性别、行业、学院等标志的具体表现为不同的类别，比如性别：男、女；行业：制造业、零售业、旅游业等；学院：商学院、法学院等。这样一类标志称为定类标志，收集的定类标志具体表现所形成的结果为定类数据。定类数据一般不能直接表现为数值资料，只能通过文字、代码和其他符号来表达，这类数据只关心类别，不关心它们的秩序，或者说排序对问题无实质性影响，所以对它们能够进行的唯一运算就是计数，即计算每一个类型中所包含总体单位的多少。若按某一标志分类后，分出的类别结果具有一定的顺序，也就是说能测定各类别的等级差异，这种标志称为定序标志。例如受教育的程度可分为大学、高中、初中、小学；考试成绩按等级分为优、良、中、及格、不及格；一个人对事物的态度分为赞成、中立、反对。这里的教育程度、成绩等级、态度就是定序标志。定序标志在个体上的具体表现所形成的结果为定序数据，可以按照某个标准排序。

定距标志不仅能将总体单位进行分类并按某个标准进行排序，而且能测定其间距的大小。如某班学生某门课程的考分不仅可以从高到低分类排序，而且可形成从最高分直到最低分的序列；这些考生的得分不仅有明确的高低之分，而且还有基本确定的测量单位，可以计算差距，例如姜伟 95 分，张利 70 分，前者比后者多出 25 分；又比如温度，今天最高温度 33℃，最低温度 18℃，相差 15℃。定距标志的具体表现所形成的结果为定距数据，可以进行加或减的运算，但不可以进行乘除的计算，其原因在于没有固定的有确定意义的零点。定比标志是比定距标志更高一级的标志，其具体表现出的数据为定比数据，不仅可以进行加减运算，而且可以作乘除运算，定比数据与定距数据的显著区别在于它有一个绝对的、非任意性的零点。例如人的年龄、体重以 0 为绝对界限，一个人的年龄不能比 0 更年轻，体重也不能比 0 更轻，因此不仅可以说 60 岁的人比 30 岁的人年长 30 岁，还可以说 60 岁人的年龄是 30 岁人的 2 倍。绝大多数的经济、管理数据都属于定比数据，如产量、产值、居民收入等。

标志按其性质可以分为品质标志和数值标志。品质标志表示事物的质的特性，是不能用数值表示的，如员工的性别、民族、工种等。数值标志表示事物的量的特性，是可以数值表示的，如员工年龄、工资、工龄等。品质标志主要用于分组，将性质不相同的总体单位划分开

来,便于计算各组的总体单位数,计算结构和比例指标。数值标志既可用于分组,也可用于计算标志总量以及其他各种指标。

标志按变异情况可分为不变标志和变异标志。当一个标志在各个单位的具体表现都相同时,这个标志称为不变标志;当一个标志在各个单位的具体表现有可能不同时,这个标志称为可变标志或变异标志。如中国第六次人口普查规定:“人口普查的对象是具有中华人民共和国国籍并在中华人民共和国国境内常住的人。”按照这一规定,在作为调查对象的人口总体中,国籍和在国境内居住是不变标志,而性别、年龄、民族、职业等则是变异标志。不变标志是构成总体的基础,因为至少必须有一个不变标志将各总体单位联结在一起,才能使它们具有“同质性”,从而构成一个总体。变异标志是数据分析研究的主要内容,因为如果标志在各总体单位之间的表现都相同,那就没有进行数据分析研究的必要了。

此外,标志还有一个名称,叫做变量,同样是用来说明事物、现象某种特征的概念。变量的具体表现为变量值。变量可以根据选择区分标准的不同,划分为定类变量、定序变量、定距变量和定比变量,或者分为品质变量和数值变量等。

### 1.2.2 模型

运用模型进行定量分析已经成为现代经济管理人员必须具备的基本功。如同高山孕育有丰富的矿产一般,数据蕴藏着事物、现象发展的规律,但需要借助模型这一工具才能开采出来。

一提到“模型”,很多人就头疼,脑海中立即浮现出了大量错综复杂的数学符号、公式,认为这是数学家们的专利,只有那些懂得高深数学理论的人才能胜任,凡人难以掌握。其实在害怕模型的同时,人们又天天在运用着各类模型。购买陕西特色小吃肉夹馍时,8元一个,买两个就得付款16元,买10个就需要花费80元,估计没有多少人算不清这笔账;出门赶路坐出租车,假设出租车计价方法为:3公里以内按起步价10元收取,超过3公里后,每公里按2元计费,若行车38公里,稍加计算,就知道需要付80元钱。如何得出的结果呢?通过简单的线性模型计算就可以完成。买东西之前,人们会在这家店先问问,到那家店再看看,货比三家之后才做决定;企业想了解生产的一批产品质量如何,通常会从中抽一些产品进行检查,然后根据检查结果,判断该批产品的合格率为多少?不合格率为多少?这就用到了稍微复杂些的抽样、参数估计模型。由此可见,人们在害怕、畏惧模型的同时,又在不知不觉中广泛地使用着。

模型在现代经济管理中的价值不言自明。如果你去问华尔街的老总们,公司中什么最值钱,他们也许会告诉你,不是那些放在地下室、存于银行的金块、珠宝,而是藏于密室的公式、算法。这一点儿也不奇怪,想象一下,国际著名的咨询公司用来进行预测的公式,公司开展知识管理在作出决策时为实现“将最合理的知识在最恰当的时间提供给最需要的人”而依赖的算法,就很容易理解了。这些模型价值连城,是公司在市场中保持垄断、战胜对手、赢得超额利润的武器,也是公司核心竞争力的集中体现。当然,这里用到的模型更加科学、严密、复杂,不然,各类管理咨询公司也不会开价那么高了。

模型是真实事物、现象的人为再现,是它所代表的真实世界对应事物、现象的抽象化结果。它略去了次要枝节,突出了主干,因而浓缩了问题的核心。模型一般具有如下特征:

(1) 模型是对事物、现象某些方面的一种描述,人们可以通过模型认识了解对象某一个

方面的性质、关系。

(2) 模型使用了专门的方式或语言,使得对事物、现象的描述变得简洁、明确。

(3) 模型是建立在一些特定假设和条件之上的,这些假设和条件限制了模型描述的范围,超出这个范围,模型就会失效。

模型在使用中,至少包含以下三个基本步骤:

(1) 建立模型。建立模型就是在充分占有各类相关数据、信息的基础上,选择一种合适的方式或语言表达事物、现象某一个方面的特征或关系。模型建立的恰当与否取决于建模者是否具备两方面的能力,一方面是模型构建知识学习的能力,另一方面是对事物、现象发展规律、趋势判断的能力。

(2) 求解模型。求解模型是在描述事物、现象一般规律的基础上,结合具体问题,在特定内外部环境、条件和时间下,求出结果。模型求解的方法包括分析法和数值法,其中,分析法是按照数学公式一步到位求出具体的解,数值法是通过某种模式逐步寻找并不断改进解的过程来求解。

(3) 解的检验。模型的解是否有效取决于模型对事物、现象发展规律近似的真实程度。利用模型得到的解有时并不一定与事物、现象的实际情况相符。这是由于所建立的模型是对实际情况的近似模拟,有的模型模拟效果可能好些,有些可能会差些;同时,在计算和推测过程中也难免会产生误差,再加上模型是在一定假设条件下建立的,所以模型的解与实际情况难免会发生偏差。因此,每次得到模型的解之后,都应对其加以分析和评价,以确定模型的解是否可信,并想出一些办法对模型的解加以修正,使之更接近于实际。此外,在条件允许的情况下,可以建立多种模型进行模拟,再经过比较或综合,确定出可信的模型的解。

在经济管理中,常用到的模型有图表模型和数学模型。图表模型就是借助各种图形和表格表达所要研究的事物、现象特征及其内部联系或与外界联系的模型,直观、醒目地显示问题的要害。图表模型一般不会提供从数量方面解决问题的具体办法,只是简单地展示影响问题的重要因素及其之间的关系。对于一个比较复杂的问题,如果借助图表作些试探,可能会有利于更清晰地建立认识逻辑。数学模型是用字母、数字和其他数学符号构成的等式或不等式。数学模型一般比较抽象,往往使人们不容易识别它所表示的问题。但是,与图表模型相比,数学模型有许多优点,如具有高度的适应性,自然科学、社会科学中各方面的问题,基本上都可以用数学模型来表示;模型建立和修改方便;容易进行运算处理;可以进行优化设计等。正因为如此,数学模型的应用比图解模型更加广泛,但就本质而言,二者是相同的,都是对实际问题的抽象描述,为人们在解决问题时抓住“牛鼻子”提供帮助。

管理人员运用模型的目的是准确预测、正确决策和恰当评估,因而,模型对管理人员而言仅仅是工具、手段。就像我们开汽车、乘火车、坐飞机去一个地方,没有必要研究汽车、火车、飞机是如何制造的一样,管理人员没有必要亲自去建立和求解模型,完全可以依靠数学家和专门的研究人员去完成。但是,运用模型解决问题不能完全由他们代劳,必须依赖管理人员与数学家既分工又密切结合地共同完成。

管理人员要明确解决什么样的问题,清楚问题的现状,为数学家提供构建模型的必要信息;任何一个模型都只是刻画事物、现象的一个或几个方面,这决定着模型都有自己的适用范围,超出这个范围,模型就会失效。因此,管理人员使用模型前需要认识模型究竟描述了事物、现象的哪个方面,模型的具体作用、特点、长处、局限以及应用中的注意事项;模型的



解的内涵是管理者必须考虑的事情，模型的解在现实中意味着什么样的情形、状况发生，会直接影响到方案的选择、决策的结果。其中，建立模型、求解模型属于数学家、专家的工作。

因此，管理人员必须始终牢记，模型是他进行预测、决策、评估的工具，他要考虑的是如何用好这个工具，而不是构造甚至操作这些工具，因此，对他们来说，最主要的是选择模型或者委托他人建立模型、求解模型，关注效益的发挥和成本的控制。形象地说，了解建立模型思路，帮助建立合适的模型就是把“货物”装到“车”上，这是管理者要动脑筋的。运算求解就是“货物运输”，只要中间不出错，外包即可，管理人员不必过多费心，而模型的解则要认真对待，发掘其中的内涵和启示。

### 1.2.3 决策

数据、模型让人们清楚认识事物、现象的内在发展规律，决策则进一步在此基础上帮助人们作出行为、方案的选择以实现其追求的目标。

决策一词出自《韩非子·孤愤》中的“智者决策於愚人，贤士程行於不肖，则贤智之士羞而人主之论悖矣”。古人亦有“凡事预则立，不预则废”的谚语传承于世。由此可见，决策很早就受到人们的重视了。

何为决策呢？从字面上来讲，就是“作出决定”，俗称拍板。其含义是在几种方案中选择其一，犹如人到多岔路口，决定要走哪一条路一样。这是对决策概念最直观、也是狭义上的一种解释。实际上任何人作出任何决定都包括明确问题或目标，提出解决问题或达到目标的各种可行方案，然后从中选择一种解决问题或达到目标的最优方案这一系列的活动过程。作出决定或拍板是指对各种方案的抉择这一活动。因此从广义上来讲，有的学者提出决策的定义为：“为实现某一特定目标，借助一定的科学手段和方法，从两个或两个以上的可行方案中选择一个最优方案，并组织实施的全部行为过程。”也有学者在定义中包括了更详细的内容，“决策是为了按预期的目的去完成某项任务或解决某个问题，运用各种方法，在系统地分析了主客观条件之后，考虑到未来的状态，根据决策准则，对提出的多种可行方案进行优选评比，选择合理方案的一种分析过程。”由于决策是人们在生活、工作中普遍存在的一种行为，因此每个人对决策都有自身的体验和直观感受。理解角度、层次的差异加上决策本身的复杂性使得决策的定义不尽相同，广泛流传的版本不下上百种。在决策众多的定义中，大多数强调了决策的目标性和对实现目标各种可能方案的选择性。应当指出的是，对目标的选择也是决策的一部分，因此决策的核心部分是对目标和为实现目标的方案的选择。由于无论采用什么样的定义也很难用简要的几句话把“决策”的定义说得完美无缺。不妨抓住“决策”概念的核心和本质，将决策定义为：对目标和为实现目标的各种方案进行抉择的过程。

决策是一个系统的过程，组成决策系统的基本要素有四个：决策目标、备选方案、自然状态和方案结果。

(1) 决策目标是决策的出发点和归宿。只有目标明确了，才便于有效地进行决策，客观地评价决策的优劣。决策目标由决策主体来抉择，取决于决策主体的偏好、信念等。有些管理人员着重于长期发展实力，而有些则强调短期的效益。决策主体在决策中发挥着重要的作用，能够控制整个决策过程。

(2) 备选方案是根据所要实现的决策目标，设计、考虑实施的方案。决策问题要求至少存在两个可供选择的方案。如果只有一种方案能够实现决策目标，那就不需要进行选择，只