



普通高等教育工商管理类专业精品教材

中央财经大学商学院 中央财经大学MBA教育中心 组编

数据、模型与决策

DATA MODELS & DECISIONS

■ 李连友 付红妍 编著





普通高等教育工商管理类专业精品教材

中央财经大学商学院 中央财经大学MBA教育中心 组编

数据、模型与决策

DATA MODELS
& DECISIONS

李连友 付红妍 编著

“数据、模型与决策”是一门关于如何收集、整理和显示数据，通过建立模型来分析、解释数据以得出结论，并利用结论进行决策的艺术和科学。本书紧紧围绕数据、模型和决策三者之间的上述逻辑关系，首先概括地介绍了收集和显示数据的方法，包括各种调查方法、表示法和图示法等；其次，重点详细介绍了通过各种不同的数学模型对数据进行分析的方法，包括代表性数据的测度与评价、概率及其分布、抽样分布与参数估计、假设检验、方差分析、相关分析与回归分析、时间序列的数据分析、指数分析等；再次，通过最优化模型简单介绍了运筹学有关线性规划和非线性规划的内容；最后，介绍了决策分析的相关内容。

本书的最大特点是紧紧围绕数据、模型与决策的思维逻辑安排内容，使读者沿着这一思路循序渐进地掌握相关知识；此外，不过于追求数学语言的完美和数学公式的推导，保证只要具备基本数学知识的学生都可以学习。

本书适合于经济与管理类院校 MBA “数据、模型与决策”和“统计学”两门课程教学使用，也适合 MPA 相关课程教学使用，还可以作为高校本科及以上“统计学”课程的教学参考书。

图书在版编目 (CIP) 数据

数据、模型与决策/李连友,付红妍编著. —北京:机械工业出版社, 2014. 1

普通高等教育工商管理类专业精品教材

ISBN 978-7-111-44510-4

I. ①数… II. ①李… ②付… III. ①决策模型 - 高等学校 - 教材
IV. ①C934

中国版本图书馆 CIP 数据核字 (2013) 第 251059 号

机械工业出版社 (北京市百万庄大街 22 号 邮政编码 100037)

策划编辑: 易敏 责任编辑: 易敏 刘静

版式设计: 霍永明 责任校对: 刘怡丹

封面设计: 鞠杨 责任印制: 李洋

三河市国英印务有限公司印刷

2014 年 1 月第 1 版第 1 次印刷

184mm × 260mm · 24 印张 · 637 千字

标准书号: ISBN 978-7-111-44510-4

定价: 49.00 元

凡购本书, 如有缺页、倒页、脱页, 由本社发行部调换

电话服务

网络服务

社服务中心: (010) 88361066 教材网: <http://www.cmpedu.com>

销售一部: (010) 68326294 机工官网: <http://www.cmpbook.com>

销售二部: (010) 88379649 机工官博: <http://weibo.com/cmp1952>

读者购书热线: (010) 88379203 封面无防伪标均为盗版

前言

“数据、模型与决策”作为一门数量分析方法方面的课程，是工商管理硕士（MBA）的必修课。本书是为适应这门课程的教学需要，由中央财经大学商学院策划编写的 MBA 课程系列教材之一，是我们对 MBA 教育的一点贡献。

尽管“数据、模型与决策”公认是“应用统计学”与“运筹学”的整合，但是从国内外已公开出版的多本教材看，不同作者对这一名称的认知不同、教学对象不同以及侧重点不同，到目前为止在教材的编排体系、所包含的内容等方面并未统一。这也充分体现了高等教育应当因材施教、百家争鸣的特点，是一种正常现象，而且，这种现象将继续保持下去。

我们认为，“数据、模型与决策”虽然是“应用统计学”与“运筹学”的整合，但是单独将“应用统计学”称为“数据、模型与决策”也未尝不可。

本书包括三部分，共 11 章。每部分及各章的主要内容如下：

第一部分（第 1 章）：数据、模型与决策概述。这一部分重点就有关数据、模型与决策的一些基本概念和三者之间的辩证关系进行了阐述，为后面的学习奠定基础。

第二部分（第 2~9 章）：主要讨论了数据收集、整理、显示和分析方法，包括：

第 2 章，数据的获取与特征描述：主要介绍了属于描述统计部分的获取一手数据和二手数据的方法；描述数据特征的表示法、图示法，以及代表性数据的测度和评价方法等。

第 3 章，概率与概率分布：主要介绍了概率的基本知识、离散型随机变量的概率分布和连续型随机变量的概率分布等。

第 4 章，抽样分布与参数估计：主要专注于各种不同抽样方法、抽样分布和总体参数估计方法的讨论。

第 5 章，参数的假设检验：主要讨论了单个总体和两个总体有关总体均值、总体比例和总体方差在各种不同情况下的假设检验问题，包括大样本和小样本、总体方差已知和总体方差未知时的各种参数假设检验的方法。

第 6 章，方差分析：主要介绍了有关方差分析的一些基本概念、单因子方差分析和双因子方差分析的方法。

第 7 章，数据的相关分析与回归分析：主要介绍了相关分析、一元线性回归和多元线性回归的基本方法等。

第 8 章，时间序列的数据分析：主要讨论了数据在时间方面的一些表现形式，各种与时间序列有关的数据特征的测度，包括各种构成要素的测度，各种水平特征、速度特征的测度，以及在此基础上对未来的预测等。

第 9 章，指数分析：主要介绍了指数的基本概念、常用的分类、各种指

数的编制理论和方法等。

第三部分（第10章和第11章）：主要讨论了最优化模型和决策分析的核心内容，属于运筹学的范畴。

第10章，最优化模型：主要就最优化模型所涉及的一些基本概念、几种最优化模型的数学建模、求解方法等基本内容进行了讨论。

第11章，决策分析：主要讨论了决策的基本问题，包括决策的含义、类型和决策步骤等；并对无概率下的决策分析、有概率下的决策分析，以及风险决策的敏感性分析等内容作了介绍。

我们深知，本书所讨论的以上内容仅仅是通行的数据、模型与决策教材中很少的一部分。作为作者，当然希望所编著的教材在体系上能够越完整越好，内容上越全面越好。但是，一个不得不考虑的重要因素是，受课时所限，不可能也不应该将大量在课堂上不讲授的内容都纳入到教材中。因此，我们在编著本书时，力求在具体内容与授课时间两方面平衡取舍，最终形成这样一种体系。

本书的特点有四：一是紧紧围绕数据、模型与决策的思维逻辑安排全书内容，使读者沿着这一思路循序渐进地掌握相关知识；二是不过于追求数学语言的完美和数学公式的推导，保证只要具备基本数学知识的学生都可以学习这门课程；三是语言通俗易懂，尽量贴近商务与经济活动管理工作的实际；四是书中大部分凡是能够用 Excel 求解的内容都予以介绍，在部分章节中还介绍了 SPSS 专业统计软件的使用方法。其中，Excel 除了第3、4、5、6章用的是2010版，第10章用的是2003版外，其余章节用的是2013版；SPSS用的是19.0版。

本书适合于经济与管理类院校 MBA “数据、模型与决策”和“统计学”两门课程教学使用，也适合 MPA 相关课程教学使用，还可以作为高校本科及以上“统计学”课程的教学参考书。

本书第1、3、4、5、6和10章由李连友教授编写，杨阳、吴蓓和周纯晓协助编写；第2、7、8、9和11章由付红妍副教授编写，侯思、何刚协助编写。最后由李连友对全书进行了审定。

本书能够编著完成并得以出版，得到了中央财经大学商学院（MBA教育中心）领导的大力支持与帮助，在此深表谢意。

自动议编著本书至最终书稿的形成，历时两年有余。在最初的编著过程中，张忠元博士一直参与了有关本书编著的指导思想、基本思路和大纲的讨论。闫霞博士对书稿进行了认真而细致的审阅。

最后，要特别指出，我们在编著本书过程中参考了大量国内外知名学者的著作，特地将这些作者的名字、著作列在书后的参考文献中。

在此，对上面所提到的所有人员表示最衷心的感谢！

李连友 付红妍
于中央财经大学

目录

前言

第1章

数据、模型与决策概述 / 1

- 1.1 工商管理中的数据 / 5
- 1.2 与数据有关的几个基本概念 / 10
- 1.3 工商管理中的模型 / 13
- 1.4 工商管理中的决策 / 13
- 1.5 数据、模型与决策的关系 / 16

第2章

数据的获取与特征描述 / 19

- 2.1 数据获取 / 20
- 2.2 数据特征描述(一)——表示法与图示法 / 23
- 2.3 数据特征描述(二)——代表性数据的测度 / 33
- 2.4 数据特征描述(三)——代表性数据的评价 / 37
- 2.5 数据特征描述(四)——数据分布的偏态与峰度 / 42
- 2.6 利用 Excel 进行数据的整理与描述 / 43

第3章

概率与概率分布 / 48

- 3.1 概率 / 49
- 3.2 离散型变量的概率分布 / 58
- 3.3 连续型变量的概率分布 / 72

第4章

抽样分布与参数估计 / 85

- 4.1 抽样 / 86
- 4.2 几种常用的抽样方式 / 89
- 4.3 抽样分布 / 93
- 4.4 参数的点估计 / 97
- 4.5 参数的区间估计 / 98
- 4.6 必要样本容量的确定 / 113

第5章

参数的假设检验 / 123

- 5.1 假设检验的基本问题 / 124
- 5.2 单个总体参数的假设检验 / 130
- 5.3 两个总体均值、比例和方差的检验 / 143

第6章

方差分析 / 159

- 6.1 方差分析的基本问题 / 160
- 6.2 单因子方差分析 / 162
- 6.3 双因子方差分析 / 171

第7章

数据的相关分析与回归分析 / 183

- 7.1 数据的相关分析 / 184
- 7.2 简单线性回归分析 / 189
- 7.3 简单线性回归分析实例 / 196
- 7.4 多元线性回归分析 / 200
- 7.5 非线性回归分析 / 205

第8章

时间序列的数据分析 / 212

- 8.1 时间序列数据 / 213
- 8.2 时间序列数据的描述分析 / 217
- 8.3 时间序列数据的成分分析 / 222
- 8.4 随机时间序列分析简介 / 252

第9章

指数分析 / 270

- 9.1 指数的基本问题 / 272
- 9.2 单一指数研究方法 / 277
- 9.3 总指数研究方法 / 278
- 9.4 指数模型分析 / 286
- 9.5 常用指数简介 / 293

第10章

最优化模型 / 299

- 10.1 约束条件下最优化的基本问题 / 300
- 10.2 线性最优化模型 / 302
- 10.3 整数线性最优化模型 / 317

10.4 非线性最优化模型 / 322

第 11 章

决策分析 / 328

11.1 决策分析的基本问题 / 329

11.2 无概率下的决策分析 / 331

11.3 有概率下的决策分析 / 337

11.4 风险决策的敏感性分析 / 342

附录

常用统计数表 / 347

附录 A 二项分布表 / 347

附录 B 泊松分布表 / 356

附录 C 标准正态分布表 / 362

附录 D t 分布表 / 363

附录 E χ^2 分布表 / 366

附录 F F 分布表 / 368

附录 G 随机数表 / 373

参考文献 / 374

CHAPTER 1

第 1 章

数据、模型与决策概述

学习目标

1. 掌握数据的概念和种类。
2. 掌握有关总体、个体、样本、变量等一些基本概念。
3. 了解模型的含义、作用和类型。
4. 掌握决策的概念、类型和步骤。
5. 了解数据、模型与决策之间的关系。

从泰坦尼克号事件中的数据谈起^①

数字工作者经常犯的一个毛病，就是常把数据当做代数问题来处理。在计算机软件多如牛毛的今天，许多人只是把所有数据往计算机软件包里一丢，再对计算机输出的一些有用没用的结果作一通胡乱解释。

须知道，数字本身是没有意义的。从事数据、模型与决策相关工作的工作者，应当从纷繁复杂的数据中，遴选出那些有用的信息，再通过建立符合客观实际的数学模型并加以求解，得出对决策有所帮助的结论，并将这一结论运用到决策实践中去。在这一过程中，数据的背景资料是非常重要的。

下面我们将结合泰坦尼克号沉船事件中的数据，谈谈如何收集、整理和描述这些数据，并运用这些数据对事故进行分析与判断。

1. 前言

泰坦尼克号（也译作铁达尼号）超豪华邮轮对每位乘客的详细资料都有记载，其中包括每位乘客的年龄、性别和所在舱位等。面对这些数据，我们不禁会提出如下一些问题：在乘客的各种变量中，哪种变量是影响生存的最重要因素？是否存在特定属性下生还概率明显不同的情况？

2. 历史背景

泰坦尼克号是19世纪末20世纪初人类科技与工艺的登峰造极之作，同时也是为了满足当时人们自信与奢华的需求而打造的。

这艘由英国白星海运公司所拥有的超豪华邮轮，由汤姆斯·安德鲁负责设计。船上装置极其堂皇，可称得上是融汇了当时的科技及精湛艺术结晶的“海上浮宫”。整条船有16个防水隔舱，双层船底，因此被认为是不沉之船。泰坦尼克号1911年5月31日下水，经过数月的试航后，1912年4月10日做首次航行，船长为E. J. 史密夫，搭载了两千多名当时社会不同阶层的乘客，于当日正午离开了南汉普顿港，预定渡过北大西洋，直达美国纽约。

3. 船体内部资料

长度：约268m（约相当于3个足球场的长度）。

甲板：9层（包括最下层甲板），分为A、B、C、D、E、F、G和G下的锅炉房。

宽度：28m。

高度：入水深60.5ft，水面至烟囱175ft（相当于11层楼高）。

乘客：头等舱329人，二等舱285人，三等舱710人。

工作人员：899人。

发动机：2个直接起动转化发动机（4个气缸，30000hp^②，75r/min），1个低压涡轮增压器（16000hp，165r/min）。

螺旋桨：3个。其中，中间规格为16ft；左右各1个，规格为23ft 6in。

总重量：46328t。

实际排水量：24900t。

锅炉：29个。

燃料：每日825t煤。

气压：215psi^③。

密室：16个，伸展至F甲板。

总运载量：3547人。

救生艇：20艘。其中，16艘木制和4艘折叠（可运载1178人）。

救生设施：3560件救生衣和49个救生圈。

① 案例改编自：《从铁达尼号沉船资料谈起》，作者林共进，收入中国统计杂志社《生活中的统计学》，由中国统计出版社于2010年出版。

② 马力，1马力=735W。

③ 磅/英寸，1psi=0.006895MPa。

全速：24n mile/h。

4. 乘客背景

其船舱分为头等舱、二等舱和三等舱。作为最高级别的头等舱，搭乘的大多为当时的贵族阶级、富翁。这些人通常会有随从跟着，如他们的仆役、护士、女婢，甚至司机等。

二等舱和三等舱的乘客大多来自不同的国家以及各种不同的社会阶级。对他们的大多数人而言，这趟旅行几乎都是他们一生中第一次搭船长途旅行的经历，甚至是第一次的旅行经历。这些人大部分都向往着美国这块新大陆，期待着能够在这块土地上再创辉煌；另外有些人则单纯是属于返乡的旅客。

5. 航线与沉没点

泰坦尼克号的处女航是从英国的南安普顿先航行到法国与爱尔兰，目的地是美国的纽约。1912年4月14日深夜，泰坦尼克号以23n mile/h的高速航行在大西洋上。深夜11点40分，当泰坦尼克号航行于北纬41°46'、西经50°14'，在新凡兰岛南部400mile处，忽然撞上游离冰山。冰山像一把锋利的钢刀在客轮左舷撕开一个大裂口，5个防水隔舱顿时破裂，船上一片混乱。到了4月15日凌晨2点20分，这艘巨轮悲哀地结束了其短促的生命，1500多名乘客葬身海底。

6. 事后发展

科学家推测，可能是冰山撞凹了约3m×9m的钢板，使铆钉因此而崩脱，海水便由接缝处灌入。但原因似乎并不如此单纯。比较钢板边缘的接缝处和船沉没前船员所看到的受损状况记录，令人觉得难以理解，而且船不到3小时就沉没了，速度也太快。1985年海洋地质学家拜勒德在深约3844m的海底找到了泰坦尼克号。令人惊讶的是，在海底找到的泰坦尼克号已断成两截，船头和船尾的距离相当远，并且到处散落着各种碎片。研究结果发现，泰坦尼克号的钢板太脆，以至于碰到冰山后，钢板没有弯曲，而是直接断裂。因为当时的人受到科学技术水平的限制，没有脆性断裂的概念，通常认为只要通过抗拉强度的测验就可以了，所以造成这一重大悲剧的发生。

7. 资料描述

泰坦尼克号具有2208笔有关生死、年龄、性别和船舱等级四种类型的资料。表1-1是各种资料的部分数据。其中，生死栏中的S表示存活，D表示死亡；年龄栏中的A表示成人，C表示小孩；性别栏中的F表示女性，M表示男性；船舱等级栏中的1、2和3分别表示头等舱、二等舱和三等舱，C表示船员。

表1-1 泰坦尼克号有关生死、年龄、性别和船舱等级部分资料表

生死	年龄	性别	船舱等级
D	A	F	1
D	A	F	1
D	A	M	1
D	A	M	1
S	A	M	1
S	A	M	1
S	A	F	1
S	A	F	1
S	A	F	1
S	A	F	1
S	A	F	1
S	A	F	1

8. 数据分析

根据上述基本资料，可以采用如下三种通常采用的数据分析方法分别对其进行分析：

(1) 条形图法。条形图法就是在一个坐标上将不同的变量用柱形固定宽度置于横轴，用柱形高度表示变量值的一种描述数据的方法。

在泰坦尼克号沉船事件的案例中，四组变量的不同条形图分别见图 1-1a、1-1b、1-1c 和 1-1d。

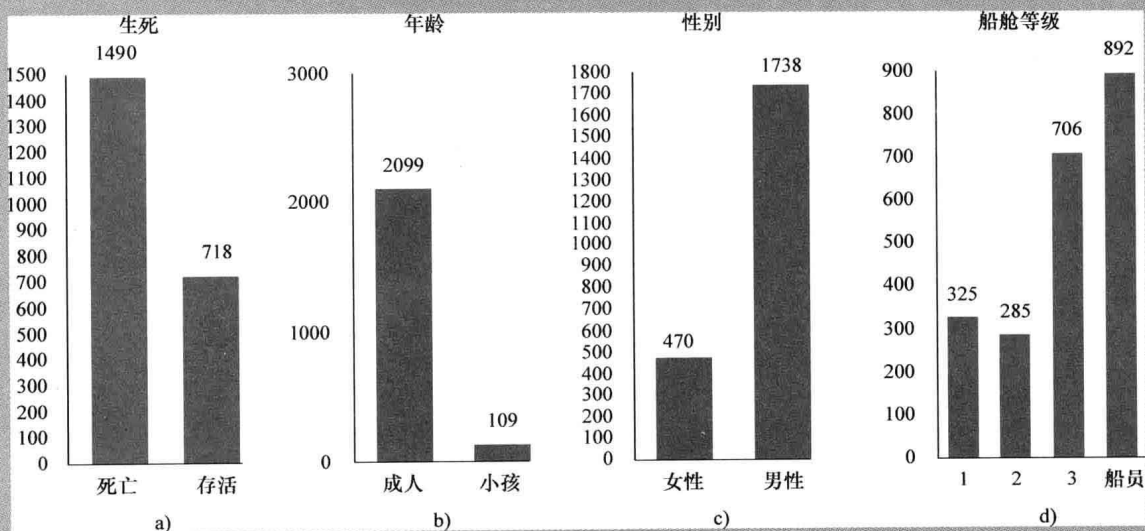


图 1-1 四组不同变量的条形图

从图 1-1 中可以大致看出各栏的分布情形：生死比率大约为 1:2；成人和小孩的比率约为 20:1；男女比率约为 3.5:1；三等舱人数比头等舱与二等舱的总和还要多。

(2) 交叉分类表法。交叉分类表是数据分析中比较常用的一种方法。它把两个变量中的一个放在横行，另一个放在纵列，通过交叉列表反映两种变量之间是否存在一定的关系。

如何判断交叉分类表中两个变量之间是否存在交互关系呢？我们以泰坦尼克号中生死与性别交叉分类表为例加以说明，如表 1-2 所示。

表 1-2 生死与性别交叉分类表

(单位：人)

	男性	女性	合计
存活	374	344	718
死亡	1364	126	1490
合计	1738	470	2208

我们知道，在全部 2208 人之中有 718 人存活，1490 人死亡，生死比率约为 1:2。试想，如果生死与性别没有关系，那么不管男性或女性他（她）们的生死比率都应该是 1:2 左右。但通过表 1-2 可以看出，在 470 名女性中有 344 人存活，126 人死亡，生死比率约为 2.7:1；同样，在 1738 名男性中有 374 人存活，1364 人死亡，生死比率约为 0.27:1。这代表着生死与性别有着密切的关系，而且可以推断，在救援过程中，优先让女性脱离危险。

其余有关生死与年龄、生死与船舱等级之间的关系也可以用交叉分类表的形式表现。

(3) 树状分类法。树状分类法是在同一总体的若干个分类中，按重要程度从高到低依次进行多重分类。其结果由于呈树状而得名。

以泰坦尼克号为例，在决定生死的诸多因素中，最重要的变量是性别。所以树状分类首先将所有数据按性别不同一分为二：男性在左，女性在右。其次，再在不同性别中分别按决定生死的最重要因素

(其中, 男性为年龄, 女性为船舱等级) 进行第二层分类。以此类推, 直至将所要分析的因素都考虑到为止。

对于泰坦尼克号中影响生死的一些因素, 绘制树状分类图如图 1-2 所示。

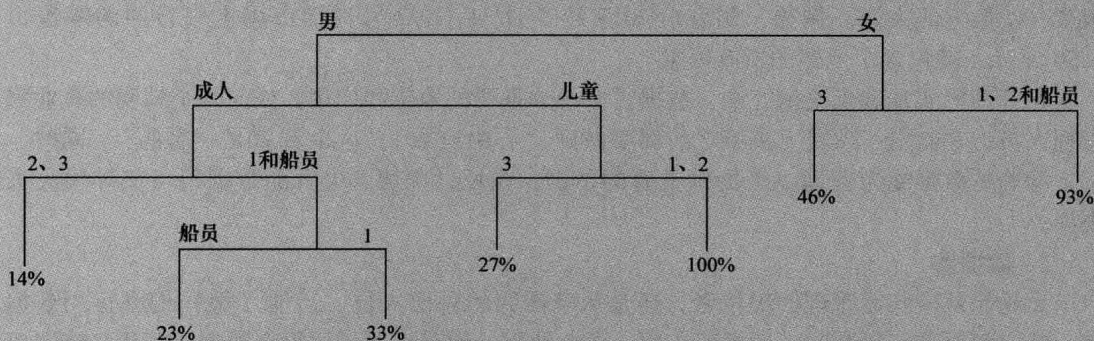


图 1-2 决定生死因素树状分类图

注: 图中的 1、2、3 分别代表头等舱、二等舱和三等舱

从上述树状分类图可以清楚地看出, 在男性方面最重要的因素是年龄。其中决定男性儿童生死的最主要因素是船舱等级: 头等舱和二等舱的男性儿童全部生还, 而三等舱的男性儿童却只有 27% 的生存率。在右面的女性资料中, 决定生死的最主要因素是船舱等级, 三等舱的生存率只有 46%, 而头等舱、二等舱和女性船员的生存率高达 93%。

总而言之, 女性与儿童的确有比较高的生存率。由此可以推断, 这应当与当时的急救措施是让妇孺优先登上救生艇有一定的关系。

另一个值得注意的是三等舱。三等舱的生存率似乎不如其他舱等高, 其中女性只有 46%, 男性儿童只有 27%。除了其他原因外, 也可以合理地推断当时船舱裂洞有可能是在三等舱。

简单的数据分析工具便可以勾勒出泰坦尼克号当时沉船的大致轮廓。当然, 数据、模型与决策中会有更多、更为复杂的数据分析方法可以使用。这里仅就几种简单的数据分析工具加以说明, 以起一个导入的作用。

在本章中, 我们以系统的方式简要介绍数据、模型和决策及其相互关系, 为后面各章详细叙述奠定基础。

1.1 工商管理中的数据

1.1.1 数据和数据集

1. 数据

数据是经过收集、整理和概括后, 用以反映某种事实的概念和数字。通俗地说, 数据就是人们通常所说的“数出有据”。正确理解数据的含义, 需要把握好以下两点:

首先, 工商管理中的数据不同于数学中的数字。工商管理中的数据是一个具体的数, 而数学中的数是一个抽象的数。要真正做到数出有据, 就必须使每个具体数据客观、真实地反映它所表达的事实在具体时间、具体空间和具体条件下的数量表现。数学中的数字则不同, 一个函数式中的每个数字可以代表任何含义。

其次，数据从大的方面看是由两个要素构成的，即反映某种事实的概念和反映该种事实状况的数字。具体说，一个完整的数据必须包括以下七个要素：①某种事实的概念或名称；②数据的计量单位；③数据的概括或计算方法；④数据的时间范畴；⑤数据的空间范畴；⑥数据的数量；⑦数据的来源。例如，北京市 2013 年 5 月 31 日 92 号汽油市场上中石油的零售价为 7.56 元/L，就包含了上述的所有因素。

在管理的实践和理论研究中，数据具有十分重要的地位和作用。任何一个精明的企业家或管理人员，在制定一项重大决策之前都要做到“心中有数”，这个数就是“数据”。同时，科学、准确的数据能为管理决策提供重要的帮助；相反，虚假、失真的数据很可能导致决策的失误。

2. 数据集

无论是从事实际管理的决策者，还是从事研究的科研人员，当他（她）们在运用数据对某一特定问题进行研究时，通常不止用一两个数据，而是需要运用很多数据。将用于特定研究而收集的所有数据的集合称为数据集。天天基金网是国内访问量比较大、用户影响较大的网站，表 1-3 是一个包含了 10 只基金信息的数据集，它们是该网站 2012 年 7 月 26 日公布的 982 只基金业绩排行榜中的一部分。

表 1-3 10 只基金业绩排行表

序号	基金代码	基金简称	最新净值	累计净值	周增长率 (%)	月增长率 (%)	本年收益率 (%)
1	000011	华夏大盘精选	7.3260	11.0060	-0.8400	-3.2500	1.5300
2	070002	嘉实增长	4.7630	5.3640	-1	0.7800	9.2200
3	070006	嘉实服务增值	3.7550	4.1350	-1.6500	0.6200	6.7700
4	162204	泰达宏利精选	3.3185	4.3985	-0.3100	1.7200	2.4300
5	340006	兴全全球视野	3.2729	3.2729	-0.8600	-0.7100	2.9100
6	110002	易方达策略成长	3.2220	4.2120	-0.8300	-0.1300	2.2300
7	260104	景顺内需增长	3.2090	4.3450	-0.8000	0.9400	-2.5800
8	510230	国泰金融 ETF	2.8660	0.8140	-2.3500	-3.1800	2.8700
9	240008	华宝收益增长	2.8279	2.8279	-0.4300	3.4000	10.8500
10	159911	鹏华民营 ETF	2.5128	0.7614	-1.4000	-4.3900	-1.1300

资料来源：天天基金网站，2012 年 7 月 26 日。

注：因表的宽度所限，本表删除了原表中日增长值和日增长率两项内容。

1.1.2 数据的测量尺度

研究者在收集和汇总数据时，应当依据由低到高的测量标准，分别按分类尺度、顺序尺度、间距尺度和比率尺度对数据进行度量，并得出以下四种不同类型的数据。

1. 分类数据

分类数据是按最低标准进行度量，只用于划分现象类别的数据。分类数据的一个例子是篮球运动员身上的号码。该号码只用来区别场上的运动员，而不是对运动员能力、价值的鉴别。也就是说，假设场上有两名运动员，其中一人的号码为 5，另外一人的号码为 10，则并不表示 10 号运动员的能力一定是 5 号的 2 倍。在商务与经济活动中，有许多数据都是分类数据。例如表 1-3 中的基金代码和基金简称等。其他常见的分类数据有人的“性别”“民族”“出生地”

“职业”等，公司的“规模”“主营业务”“信誉程度”“所属行业”等。

2. 顺序数据

顺序数据是按次低标准度量，除了具有分类数据的功能外，还可以为所研究对象进行排序的数据。例如，表 1-3 中的“序号”就是一个顺序数据，它按基金的最新净值从高到低排序。

从表 1-3 的例子中还可以看出，利用顺序数据只能为所研究的对象排序，而不能确定各序数之间的相等区间，即排在第 1 位的华夏大盘精选与排在第 2 位的嘉实增长，以及排在第 2 位的嘉实增长与排在第 3 位的嘉实服务增值之间的相等区间（后面类推）。

有些研究者在问卷调查的备选答案中就常使用顺序数据的形式。例如，银行在对客户进行有关工作人员服务满意度调查方面经常设计的备选答案是：

您对工作人员的服务态度感到 _____
 (请在横线上打√) 满意 比较满意 不满意

在上述三个选项中，所有人都会同意“满意”要比“比较满意”好，而“比较满意”又要比“不满意”好，但几乎所有受访问者都不可能认为二者之间的差是相等的。

3. 间距数据

比分类数据和顺序数据更高一个类别的数据是间距数据。间距数据是一种除了具有顺序数据的所有特征外，给出数据之间具体距离的数据。间距数据是对顺序数据间的距离的测度。这种测度通常使用自然的或物理的单位作为计量尺度，如人口用“人”来度量、温度用摄氏的或华氏的“度”来度量、重量用“克”来度量等。

间距数据表现为连续数据时，其之间的差所代表的距离是相等的。这方面的一个例子是使用摄氏温度对气温进行度量，连续数字如 21℃、22℃、23℃……之间的温度差数是一样的。同时，由于连续数据之间的差是相等的，因此，只要给出一个度量单位，就可以准确地指出两个计数之间的差值。例如，一个地区年平均气温是 15℃，而另一个地区的年平均气温是 20℃，就可以容易地指出两个地区年平均气温相差 5℃。

间距数据的另一个特点是，两个数值之间只能计算差值，而它们的比值没有意义。比如，在上面有关两个地区之间气温差的例子中，可以说两个地区年平均气温相差 5℃，却不能说后一个地区的年平均气温是前一个地区的 1.33 倍。

需要指出的是，在使用间距数据时，零点的选定只是为了方便或惯例，而不是通常所指的自然数或固定的零点。也就是说，零只是坐标上的一个点，并不表示该现象没有或不存在。这方面最常见的一个例子就是，当某天的气温为 0℃ 时并不是说这天就没有温度。又如，一个学生数学的考试成绩为 0 分，并不能代表这名学生没有数学成绩或没有任何数学知识。

4. 比率数据

比率数据是一种除了具有间距数据某些特征之外，有绝对零点并且两个数字间的比值具有意义的数据。它是最高度量标准的数据。比率数据的一个例子是当公司一名员工的月收入是 3000 元，而得知另一名员工的月收入是 1500 元时，则不仅可以说前者月收入比后者多 1500 元，还可以说前者的月收入是后者的 2 倍。也就是说，在这个例子中，不仅数据之差有意义，其比值也具有实际意义。

比率数据中的绝对零点是指零是固定的，代表不具有所研究特征的数据的值。比如，当一名员工与他受聘的公司解除劳务合同关系时，他从该公司得到的收入就为 0。换句话说，该员工就不具有所研究特征数据（收入）的值。此外，由于零点是代表一个固定数值，因此它不能被任意指定。

1.1.3 各种不同度量标准数据的比较

上述各种不同数据之间既有联系也存在着差别。首先，从度量标准的级别看，图 1-3 给出了从低到高的关系图。

其次，从各种数据的使用范围看，图 1-4 给出了四种类型数据使用范围的关系。

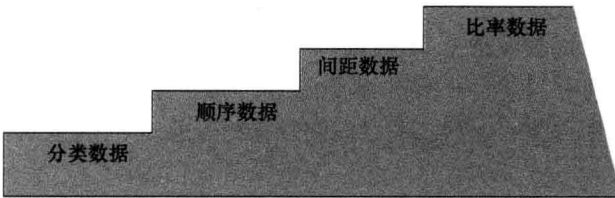


图 1-3 不同度量标准数据层次关系图

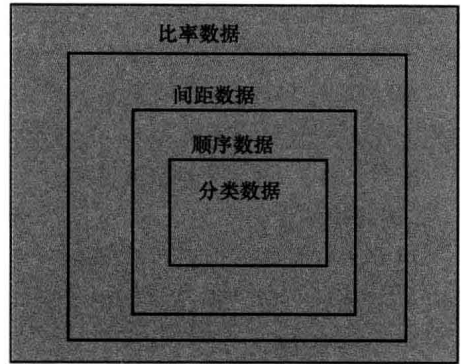


图 1-4 四种类型数据使用范围的关系图

图 1-4 中的同心正方形表示高级别的数据可以使用低级别的统计方法进行分析，而低级别的数据不能使用高级别的统计方法进行分析。例如，对于分类数据和顺序数据，尽管两种数据都不能进行任何数学运算，但顺序数据除了具有分类数据可以将研究对象区分为不同类别的功能外，还可以比较不同计量结果的好坏、大小等，其统计分析方法的使用范围比分类数据广泛一些。而对于第三个层次的间距数据来说，除了可以对研究对象进行分类和排序以外，还可以通过加减数学运算确定出不同顺序间的准确差值。显然，它的使用范围要比分类数据和顺序数据都广泛，但仍不能通过乘除运算来确定数值间的比值。作为最高级别的比率数据则具备了数据的所有功能，即除了可以将研究对象区分为不同类别、对不同类别进行排序和计算出各类别数据之间的差值外，还可以通过乘除的数学运算来确定数值间的比值。因此，比率数据是应用范围最广泛的数据。

1.1.4 数据的其他类型

数据除了可以按不同测量尺度区分为分类数据、顺序数据、间距数据和比率数据外，还可以按其他不同的标准对其进行分类。在工商管理中，有用的数据分类还有以下几种。

1. 质量数据和数量数据

质量数据和数量数据是按照数据所反映现象（或事物）的不同特征划分的，也是对不同测量尺度数据的进一步概括。

质量数据是指为某一现象（或事物）的性质提供标记或名称的数据。表 1-4 是某公司一项关于员工对某项政策观点的调查数据表，其中可以看到被访者的工作证号、性别、受教育程度和观点等都是质量数据。质量数据包括分类数据和顺序数据。

由于质量数据都来自不太精确的度量，因此，在分析使用这类数据时，只需了解总体中每种类型个体的标记、顺序以及各自的分布频数或比率就够了。另外，有时即使质量数据用数字代码表示，但对其进行加、减、乘、除等数学运算也是没有意义的。

表 1-4 某公司员工对某项政策观点的调查数据表

员工工作证号	性别	受教育程度	观点	月收入/元
001	男	高	支持	6000
002	男	中	说不准	4850
003	女	高	支持	5760
004	男	高	支持	5700
005	女	低	反对	2930
⋮	⋮	⋮	⋮	⋮
499	男	中	支持	3900
500	男	低	反对	3240

数量数据是指明现象（或事物）的数量为多少，或者有多少个的数据，包括间距数据和比率数据。表 1-4 中月收入的数据就是数量数据。

对于数量数据进行数学运算可以得到满足分析和决策需要的有意义的结果。例如，对于表 1-4 中 500 名员工的月收入，可以先求和观察他们的月总收入是多少；然后再通过计算其均值反映他们月收入的一般水平。

2. 横截面数据、时间序列数据和面板数据

在工商管理活动中，为更好地分析、使用各种数据，还可以按时空要素将数据区分为横截面数据、时间序列数据和面板数据。

横截面数据是在同一时点或是在大约在同一时点所收集到的不同空间的数据。例如，表 1-3 中的数据是横截面数据，因为它们描述的是 10 只基金 8 个变量在同一时间上的状况。

时间序列数据是在若干个时期或时点上收集到的数据。其中，如果数据所反映的为现象在一段时期内的数量表现，则这类数据称为时期数据。例如，表 1-5 中的数据就是一些时期数据。它们反映了万科 A（000002）2007～2011 年这五年间历年主营业务利润的状况。

表 1-5 万科 A（000002）2007～2011 年主营业务利润的状况

年 份	2007	2008	2009	2010	2011
主营业务利润/ 万元	1080349.93	1145263.18	1076371.51	1501624.74	2077580.01

资料来源：新浪财经网。

如果数据所反映的为现象在某一时点的数量表现，则这类数据称为时点数据。图 1-5 就是一个由时点数据构成的时间序列。该时间序列中的每个数据都反映了北京当日市场上 93 号汽油（现为 92 号）的价格。

为了便于收集和应用好时间序列数据，有必要对时期数据和时点数据进行区分。二者主要的区别在于：对不同时间的时期数据进行加总可以得出有意义的结果；时点数据则不同，将不同时点的数据加总在一起是没有意义的。

面板数据是将横截面数据和时间序列数据综合起来的一种数据。由于面板数据有时间序列和截面两个维度，当它按两个维度排列时，整个表格像是一个面板。例如，表 1-6 有关华北五省、市、区 2005～2011 年的货物进出口总额情况就是一张反映面板数据的表格。