

高等院校工商管理专业规划教材

The Essence of Managerial Statistics

简明管理统计

贾怀勤 编著



对外经济贸易大学出版社

University of International Business and Economics Press

高等院校工商管理专业规划教材

简明管理统计

贾怀勤 编著

对外经济贸易大学出版社

图书在版编目 (CIP) 数据

简明管理统计/贾怀勤编著. —北京: 对外经济贸易大学出版社, 2007

高等院校工商管理专业规划教材

ISBN 978-7-81134-017-4

I. 简… II. 贾… III. 经济统计学 - 高等学校 - 教材
IV. F222

中国版本图书馆 CIP 数据核字 (2007) 第 178208 号

© 2007 年 对外经济贸易大学出版社出版发行

版权所有 翻印必究

简明管理统计

贾怀勤 编著

责任编辑: 宋志红

对外经济贸易大学出版社

北京市朝阳区惠新东街 10 号 邮政编码: 100029

邮购电话: 010-64492338 发行部电话: 010-64492342

网址: <http://www.uibep.com> E-mail: uibep@126.com

山东省沂南县汇丰印刷有限公司印装 新华书店北京发行所发行

成品尺寸: 185mm × 230mm 15.25 印张 305 千字

2007 年 12 月北京第 1 版 2007 年 12 月第 1 次印刷

ISBN 978-7-81134-017-4

印数: 0 001 - 5 000 册 定价: 25.00 元

出版说明

改革开放二十年来,中国经济得到了稳定、迅速和健康的发展。处于转型经济时期的中国,在迎接全球化挑战的同时,也面临着前所未有的发展机遇。学术界也是如此。在与国外学者进行学术交流与对话的同时,中国学者也逐渐适应了国外规范的研究模式。他们开始接受一些在早期的研究中不被重视的基本概念:构念、量表、调研、信度、效度、实证方法,等等。幸运的是,处于转型经济背景下的中国学者能够更为直接地感受到中国的企业管理正在发生的变化,这使他们能够比其国外同行在研究中国情境下的问题时有更多的发言权。

工商管理是一门实践性、应用性很强的学科。教学之中心在于教材。1978年的改革开放政策最终导致了20世纪90年代初的教材引进狂热。其后的十余年,国外原版教材、影印教材、翻译版教材几乎涉及了各类自然科学和社会科学学科。从工商管理类教材来看,现在,中国的学者们开始逐渐意识到:国外教材是特定情境下的产物,在很多方面与中国的国情并不相符,决不能不加取舍地照搬。因此,以“全球化思维、本土化实践”来编写教材已成为中青年学者责无旁贷之任务。

教材不仅要传播既有的知识,而且要反映学科的前沿动态。依托于国内最早(1983年)引进西方工商管理教育体系、首家(1988年)获得美国同类大学MBA资格承认的中国商学院——对外经济贸易大学国际商学院,对外经济贸易大学出版社拥有足够的实力和能力号召兄弟院校中的具有丰富教学和研究经验的教师加入作者队伍。

本次组织编写的“高等院校工商管理专业规划教材”,参考了国际上通行的管理专业核心课程的设置,按由浅入深、由一般到特殊的原则,充分考虑了我国工商管理专业课程的设置要求,涵盖了工商管理专业的全部必修课程。

本套教材的开发始终坚持高水平和高标准的原则。教材的适用面较广,既可适应高等院校工商管理专业及不同层次教学的需要,又可供各类企业管理人员培训以及个人自学使用。

我们希望本套教材的推出能够为中国工商管理教育的发展作出积极的贡献。如果您在教材使用过程中发现任何问题或改进意见,欢迎您与我们联系,以便我们在后期不断进行完善!

对外经济贸易大学出版社
2007年5月

前 言

本书定名为“简明管理统计”，力求适应近年来管理类研究生统计学教学的发展变化。

我国管理类研究生的统计学教学经历了由综合到分解的过程，先是力图将经济和管理所需用的统计知识和方法尽可能地装进一门统计课程中，后来逐渐平行地开设出多门统计分支学科课程，而把基干的统计知识和方法保留在“管理统计”中。尽管如此，“管理统计”教科书的编者还是在书中尽可能多地包括管理统计的各方面知识和各种方法，于是造成各相关课程教学内容的不适当重复。本书的编写，就是考虑到学生在修习“管理统计”之后还将修习“商务预测”、“多元统计分析”和“计量经济学”等课程，将“管理统计”的知识点限定于基干的统计知识和方法，不与后续课程的教科书重复。此乃“简”字的第一个涵义。

近年来统计学教科书的另一个趋势是，随着一些大型专门统计分析软件的引进，教科书在阐述统计分析方法时，常常附带列出软件程序语言或视窗对话框影像，为的是方便教师讲授和学生演练，以至于软件操作介绍的文字和影像占据了相当大篇幅。本书作者认为，由于这些大型专门统计分析软件工具书的编印、发行和操作知识的逐渐普及，在管理统计教科书中实无必要再详列操作指南内容。因此，本书只将有关分析结果打印输出，以增强学生对统计方法的理解，而对软件工具只给出其进入路径。此乃“简”字的第二个涵义。

管理类研究生在其本科阶段有的学过统计学相关课程，有的则根本没有系统地接触过统计学知识。为此，这门课程不得不从统计基础知识讲起，但又不能完全重复本科统计学的知识点和套路。如何使后一种背景的研究生在有限的学时内掌握由统计描述到统计推断的基干知识和方法，同时又使前一种背景的研究生感到修习此课程有新的收获，在教材的知识展开框架上必须有双面适应的独到之处。对此，本书：（1）在知识展开层次上依“经验数据的统计描述”——“数据的理论分布模型”——“分布特征的统计推断”线索渐次展开；（2）在统计推断方法上按定标数据、定秩数据和定名数据三个级次分别演绎；（3）在数据分布基本趋势研究上遵循先集中后离散的顺序；（4）在变量数目上从单一变量到双变量再到多变量逐步推进。为全书搭建这样一个架构，无非是想避免学生接受了一大串统计方法而不辨其应用对象和场合的乱用以及统计分析软件功能的滥用。此乃“明”字的涵义。

书中转引和改编的例题尽可能注明出处。另外一些例题则是在近年间 MBA 学生提供的案例素材基础上编写的。在此谨向有关资料提供者致谢。

书中所涉及的部分数据，可在对外经济贸易大学出版社网站上获取，网址是：
www.uibep.com。

对于书中可能存在的不妥和疏漏之处，诚恳欢迎同行和读者不吝指正。



2007 年 8 月于惠园

目 录

Contents

第 1 章 经验数据的描述方法 (上)	(1)
1.1 统计描述与统计推断	(1)
1.2 数据级次和分类	(4)
1.3 数据整理和分布特征展现	(5)
1.4 交叉分组	(13)
第 2 章 经验数据的描述方法 (下)	(22)
2.1 众数与异众比率	(22)
2.2 五大位次量数	(24)
2.3 均值与方差	(25)
2.4 偏态系数和峰度系数	(27)
2.5 探索性工具	(29)
第 3 章 数据的理论分布模型	(36)
3.1 总体数据理论分布的概念和基本类型	(36)
3.2 概率分布参数: 期望值和方差	(39)
3.3 几种简单的理论分布模型	(40)
3.4 二项分布	(42)
3.5 超几何分布	(43)
3.6 正态分布	(45)
3.7 χ^2 分布、 F 分布和 t 分布	(47)
第 4 章 单一总体定标数据集中趋势量数的估计	(54)
4.1 参数与估计量	(54)
4.2 样本均值的抽样分布理论模型	(56)
4.3 总体均值区间估计的基本原理	(57)
4.4 σ^2 已知时 μ 的区间估计	(58)
4.5 σ^2 未知时 μ 的区间估计	(58)
4.6 必要样本容量的确定	(59)

第5章 单一总体定标数据集中趋势量数的假设检验	(62)
5.1 假设检验的基本原理	(62)
5.2 有关假设检验的基本概念	(63)
5.3 总体标准差明确条件下均值的假设检验	(67)
5.4 总体标准差不明确条件下均值的假设检验	(68)
5.5 P 值检验法	(69)
5.6 β 值的计算和必要样本量的确定	(71)
第6章 单一总体定标数据离散趋势量数的估计和假设检验	(76)
6.1 方差研究的重要性和方差的抽样分布	(76)
6.2 σ^2 的区间估计	(77)
6.3 σ^2 的假设检验	(78)
6.4 σ^2 统计推断中的大样本正态近似法	(79)
第7章 两总体定标数据参数间差异的估计和假设检验	(83)
7.1 两总体均值差量的统计推断 (独立样本)	(83)
7.2 两总体均值差量的统计推断 (配对样本)	(88)
7.3 两总体方差差异的统计推断	(90)
第8章 多总体定标数据集中趋势量数间差异的统计推断	(96)
8.1 方差分析的作用和概念	(96)
8.2 单因素方差分析	(97)
8.3 双因素方差分析	(104)
第9章 定名数据的统计推断	(116)
9.1 单一总体比率的统计推断	(117)
9.2 两总体比率间差异的统计推断	(119)
9.3 单一总体多值分类比率的假设检验	(123)
9.4 多总体多值分类齐一性的假设检验 (独立样本)	(125)
9.5 多总体双值分类齐一性的假设检验 (关联样本)	(128)
9.6 拟合优度检验	(130)
9.7 游程检验	(135)
第10章 定秩数据的统计推断	(140)
10.1 非参数统计	(140)
10.2 单一总体位势检验——符号检验	(140)
10.3 两总体位势差异性检验 (配对样本) ——符号检验	(144)
10.4 两总体位势差异性检验 (配对样本) ——威尔柯克松符号—秩 检验	(145)

10.5	两总体位势差异性检验 (独立样本) —— 符号秩和检验	(147)
10.6	多总体位势差异性检验 (独立样本) —— 克鲁斯卡尔—华利思 一向方差分析	(150)
10.7	多总体位势差异性检验 (关联样本) —— 弗里德曼二向方差分析 ...	(151)
第 11 章	关联分析	(158)
11.1	定标数据的关联分析	(158)
11.2	定秩数据的关联分析	(164)
11.3	定名数据的关联分析	(169)
第 12 章	回归分析	(177)
12.1	线性回归的总体模型	(177)
12.2	线性回归的统计描述	(178)
12.3	回归模型的统计推断	(185)
12.4	线性回归的扩展	(193)
12.5	回归建模的若干问题	(200)
统计数表		(216)
附表 1	二项分布累积概率表	(216)
附表 2	标准正态分布表	(219)
附表 3	卡方 (χ^2) 分布表	(221)
附表 4	F 分布表	(222)
附表 5	t 分布表	(229)
附表 6	塔基的学生氏化极差 q 值表	(231)
附表 7	杜宾—沃岑检验上下界 (DW 统计量临界值)	(232)

第 1 章

经验数据的描述方法（上）

本章学习目标 ■

学完本章，应该明了数据的总框架，掌握对数据进行整理和表述的基本方法。

1.1 统计描述与统计推断

管理，无论是宏观管理还是微观管理，无论是企业管理还是公共管理，都需要借助于数据对所管理的对象的存在方式和变化趋势作出认识、判断和推测。管理统计就是使用统计手段对管理所需数据进行采集和加工，以导出对管理决策有支持作用的信息的一整套科学方法。管理者面对的是管理对象的全体，如一级行政区政府面对的是本行政区的全体居民，企业营销策划部面对的是本企业产品现实和潜在用户全体。因此，其必须掌握能够概括管理对象全体的数据特征的信息，方能作出科学决策。

就一个管理课题而言，承担其研究内容的全部事物或现象的集合，即构成统计总体，简称总体。而该集合中的每一个单元，就称为总体单位。含有无数多个总体单位的总体称为无限总体，否则即是有限总体。总体单位在各方面的属性称为标志，如企业的行业、规模、所有制类别、赢利状态，等等。对某个统计总体单位一标志的测量获得其标志值。多个总体单位同一标志的标志值集合在一起，就构成数据。当对数据进行定量分析时，该属性标志即称为变量，而构成数据的每个标志值，即是变量值。

对于总体的研究，按所接触的总体单位的全面性与否，存在两种组织方式。一个是普查，即采集全部总体单位的数据，直接认识总体。另一个是抽样调查，即采集部分总体单位的数据，通过部分来认识总体。后者中，部分总体单位的集合，即构成样本。从总体中选取这部分单位的过程称为抽样。如果遵循随机原则从总体中抽取样本，就是随机抽样；否则就是非随机抽样。随机抽样产生的样本有助于对总体数量特征作出较为准

确和有把握的认识。

无论是普查还是抽样调查，它们都上从所研究的对象的数量表现出发，经过一系列统计操作，达到认识总体数量特征的目的。只不过后者经过了更为复杂的操作过程。

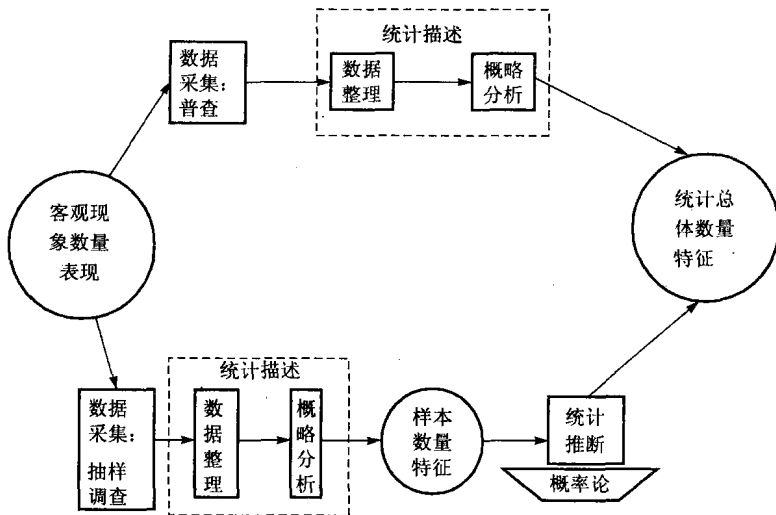


图 1-1 统计描述和统计推断的路径

就采集到手的数据本身进行整理，并对其数量特征进行概略分析，这个操作程序即所谓统计描述。统计描述作用于普查数据，可达到直接认识总体数量特征的目的。而统计描述作用于样本数据后，直接认识的只是样本数量特征；还必须在概率分布模型理论支持下，从样本数量特征出发，进一步推及总体数量特征，这就是所谓统计推断。从数据加工流程看，似乎抽样调查更为繁琐，但是就整个管理研究看，抽样调查较普查有更多的适用性。第一方面原因是，对于无限总体无法进行普查，即使对有限总体，也无法承受耗竭性普查。第二方面原因是，普查涉及单位非常之多，操作起来非常复杂，要避免操作误差得付出非常大的代价，而抽样调查成本相对较低，比较容易避免操作误差，对随机误差也可以进行事先推算和控制。

【例 1-1】 加拿大公众节能态度研究^①

1973 年，第一次全球能源危机引起了西方发达国家对能源问题的战略认识，从能

^① 加拿大 Wilfrid Laurier 大学 Dr. Gerald Keller 和 Dr. Gordon H. G. McDougall 1984 年 7 月提交给 Policy Co-ordination Bureau, Consumer and Corporate Affairs Canada 的研究报告《加拿大公众对于能源形势态度的调查》。Dr. Gerald Keller 是本书作者的管理统计学指导教师。

源战略保障、石油能源替代、能源节约等各方面开展了研究。加拿大某研究机构自1975年以来持续开展公众对于能源形势态度的研究。1979~1984年在数个中心城市使用电话开展随机抽样调查,具体的样本构成见表1-1。

表1-1 1975~1984加拿大公众节能态度调查样本构成

城市	1975	1976	1977	1978	1979	1980	1981	1982	1983	1984
哈利法克斯	250	250	270	251	206	200	201	200	208	200
蒙克顿	—	—	—	—	—	—	—	—	210	200
魁北克城	293	288	281	275	205	207	200	207	212	201
蒙特利尔	278	288	277	276	214	205	200	201	213	201
渥太华	—	—	—	—	206	205	200	213	210	201
多伦多	251	262	260	251	206	205	203	202	213	202
温尼伯	249	253	237	250	205	204	201	205	212	203
卡尔加里	249	250	246	255	206	205	201	224	210	204
温哥华	251	249	244	250	206	206	200	207	209	202
合计	1821	1840	1815	1808	1654	1637	1606	1659	1897	1814

资料来源: *A Survey of the Canadian Public's Attitudes Towards the Energy Situation, Volume II.*

以下是1983年和1984年的问卷。

节能研究问卷

1. 您认为下述两个问题对我们生活的影响程度如何?

(a) 能源短缺

很严重 ☐

有些严重 ☐

不甚严重 ☐

根本不严重 ☐

(b) 能源价格

很严重 ☐

有些严重 ☐

不甚严重 ☐

根本不严重 ☐

2. 您认为在过去12个月中汽油价格上升了百分之几? _____ (精确到5%整数倍)

3. 能源新形势出现以来,您是否采取了下述节能措施?

(a) 降低采暖温度

☐是

☐否

(b) 注意随手关灯

☐是

☐否

(c) 少驾车出行

☐是

☐否

(d) 尽量乘公共汽车

☐是

☐否

4. 您家里若有恒温取暖装置,

- (a) 其日间温度设定在_____°F。
 (b) 其夜间温度设定在_____°F。
5. 您是否同意下述关于能源形势的提法?
- (a) “能源危机”纯系一种虚构。 ☐是 ☐否
 (b) 您对公用事业当局和政府号召人们更多地节能感到厌烦。 ☐是 ☐否
 (c) 石油公司利用当前能源形势捞到了不少好处。 ☐是 ☐否
6. 您的年龄是_____岁。
7. 您的性别是 ☐1 = 男 ☐2 = 女
8. 您受过_____年正规教育。
9. 去年您全家的全年收入是_____美元 (按 5 000 美元的整数倍计)。

1.2 数据级次和分类

数据是对所研究事物、现象的属性的定性的具体描述。从不同的角度看问题,数据就有不同的分类方法。管理统计将数据按其所蕴涵的内容深度分为 4 个等级:定名数据、定秩数据、定距数据和定比数据。

定名数据是仅对事物、现象所属类别的具体描述。它们不表示顺序,没有多少大小之分。例如例 1-1 的问项 7 调查的是性别,用 1 代表男性,用 2 代表女性。也可以反过来,用 2 代表男性,用 1 代表女性。

定秩数据,是对事物、现象所具有的属性顺序的具体描述。它们描述属性的强弱,测定结果按档次排列,然而各档次之间的间隔是不确定的。对个体进行测定得一序列尺度数码,不仅表明该个体与具有相同数码的个体属性程度相同,还表明它与具有与之不同数码的其他个体属性程度的不同。例如例 1-1 的问项 1 (a) 调查对如何看待能源短缺的影响程度,分别有“很严重”、“有些严重”、“不甚严重”和“根本不严重”4 种答案供选择,它们是依次减弱的。当然也可以翻转过来,按依次增强顺序排列。但是不可以随意交换答案的位置。

定距数据的描述功能高于定秩数据,它们不仅能表明事物、现象的强弱程度,而且以确定的间距表明这种强弱程度。定距数据的数值间有确定的间距,但是没有自然确定的原点,换言之,它的零点是人为指定的。因此,计算定距数据之间的比例没有意义。温度是典型的间距尺度。例 1-1 的问项 4 调查的取暖装置日间和夜间设定的温度,就是采集定距数据。经济和工商管理中常见的定距数据的例子还有统计决策中的效用 (Utility) 和统计指数。

定比数据是最高级的数据，它不仅有确定的间距，而且有自然确定的零点。例1-1的问项6、8和9分别调查年龄、受教育年数和全年收入，都是采集定比数据。

四种数据的描述功能是依次增大的，相应地，它们的可计算性也是依次增大的。由于使用定距数据的机会较少，而且在许多场合可以同定比数据施以同样的处理和分析方法，在不需要特别区别时往往把两者归成一类，称为定标数据。

传统的统计学使用两分法：质别数据和量别数据。质别数据对应于定名数据和定秩数据，量别数据对应于定标数据。

数据还有离散型和连续型之分。离散型数据只能取特定的数值（一般为整数），连续型数据可以取介于两个特定值之间的任何数值。定标数据可以是连续型的，也可以是离散型的。定名数据和定秩数据只能是离散型的。

在同一时期（或时点）对客观总合现象大量观察所得数据，称为截面数据。截面数据按一定规则排列，形成截面数列。相对于截面数据，由反映同一总体数量特征的数据进行历史积累，可得到一系列的数据，将其按时间顺序排列，即形成时间数列。在同一时间顺序上平行排列着两组或两组以上的时间系列数据，就构成了面板数据。本书只研究截面数据问题。

1.3 数据整理和分布特征展现

从现场采集的样本数据或总体数据是原始数据，它们呈散在状态，杂乱无章，很难看清楚这些数据的联系，也不能展示出样本或总体的数据特征。因此必须对原始数据进行整理。

表 1-2 是例 1-1 的数据采集结果片断（全部数据另有 EXCEL 文件提供）。为便于举例，将样本压缩，1983 年和 1984 年各取 60 个家庭。

表 1-2 能源研究調查數據 (片段)

年 份	编 号	1	1	2	3	3	3	3	4	4	5	5	5⑥	6	7	8	9
		(a)	(b)		(a)	(b)	(c)	(d)	(a)	(b)	(a)	(b)					
		短 缺	价 格	涨 价	降 温	关 灯	少 驾	公 交	日 温	夜 温	虚 构	厌 倦	好 处	年 龄	性 别	教 育	收 入
1983	1	2	2	30	1	1	1	0	60	60	1	1	0	37	2	15	40
1983	2	1	3	10	1	0	0	1	60	60	1	0	0	25	2	5	20
1983	3	1	3	10	1	0	1	0	65	65	1	0	0	61	2	16	40

续表

年 份	编 号	1 (a)	1 (b)	2	3 (a)	3 (b)	3 (c)	3 (d)	4 (a)	4 (b)	5 (a)	5 (b)	5 ^⑥	6	7	8	9
		短缺	价格	涨价	降温	关灯	少驾	公交	日温	夜温	虚构	厌倦	好处	年龄	性别	教育	收入
1983	60	1	3	10	1	1	0	0	70	70	0	0	0	30	1	12	40
1984	1	1	3	5	0	0	0	1	70	70	0	0	0	40	1	15	25
.....																	
1984	60	1	3	40	1	1	1	1	70	70	1	0	0	60	1	11	20

资料来源：同表 1-1。

欲观察数据的条理，应进行分组，编制频数分布表。所谓分组，对于质别数据就是依据属性的不同将数据划分成若干组，对于量别数据就是依据属性数值的不同将数据划分成若干组。在此基础上，将个体数据分配到各组之中，计数落入各组的个体数目——称其为频数，就形成了一组数据的频数分布。

频数分布的具体展开形式就是频数分布表，表的基本框架是两列，左为数据的分组标写，右为各组的频数。组别的划定应遵循穷尽和互斥的原则，即一个数据必须能分配进入一个特定的组，并且一个数据只能分配进入一个特定的组。频数一列的右侧还可以在派生出其他列：质别分组可以派生出相对频数——即各组频数占数据总数目的百分比；量别分组，除派生相对频数外，还可以有累计频数和相对累计频数。

1.3.1 质别数据分组汇总和相应的统计图

根据质别数据的属性不同进行分组，再计数归入各组的频数。表 1-3 是定名数据——1983 年能源研究调查对象的性别的分组和汇总结果。表 1-4 是定秩数据——调查对象对能源价格影响程度的判断的分组和汇总结果，按 1983 年和 1984 年平行列出，便于比较。

表 1-3 1983 年能源研究调查对象的性别

性别	人数	相对频数 (%)
男	30	50
女	30	50
合计	60	100

资料来源：同表 1-1。

表 1-4 调查对象对能源价格影响程度的判断的频数分布

程度	1983 年	1984 年
1 = 很严重	6	11
2 = 有些严重	15	14
3 = 不甚严重	29	21
4 = 根本不严重	10	14
合计	60	60

资料来源：同表 1-1。

通常用饼图和条形图展示质别数据的频数分布态势。例如可以用饼图展示调查对象对能源价格影响程度的判断分布（见图 1-2），还可以用饼图的扩展形式—园环图展示 1983 年和 1984 年调查对象对能源价格影响程度的判断分布，以便进行对比（见图 1-3）。图 1-2 标注的是频数；图 1-3 标注的是相对频数，1983 年在内环，1984 年在外环。

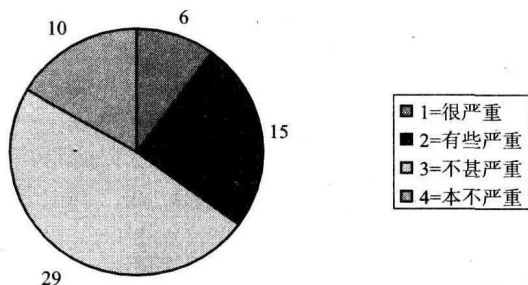


图 1-2 对价格影响程度不同判断的分布

资料来源：同表 1-1。

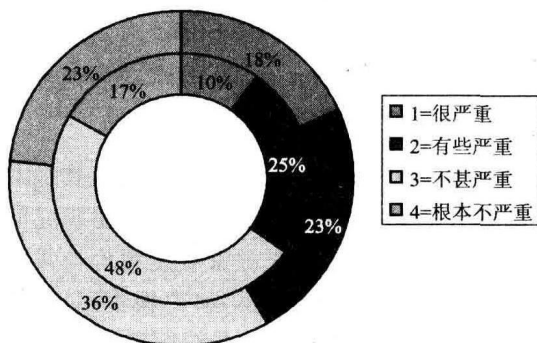


图 1-3 1983 年和 1984 年对价格影响程度的判断

资料来源：同表 1-1。

只分为两组的数据一般不绘制饼图，但可以绘制条形图（见图 1-4）。对于需要对比的数据可以绘制并列条形图（见图 1-5）。

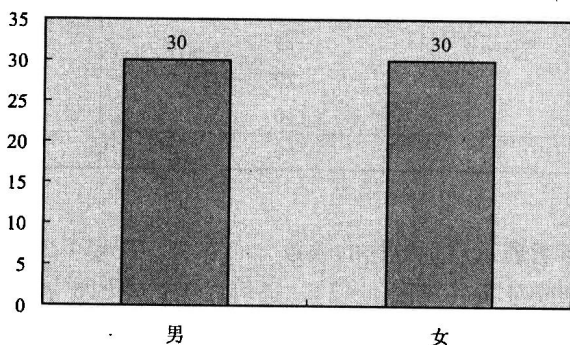


图 1-4 调查对象的性别分布

资料来源：同表 1-1。

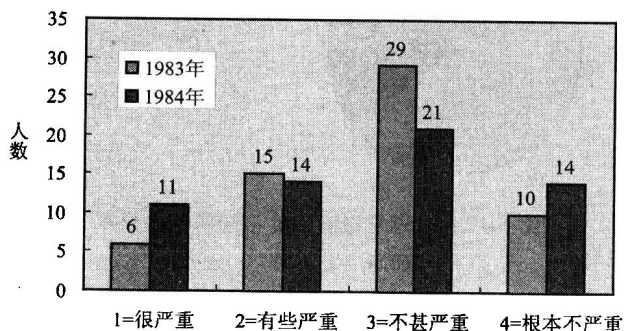


图 1-5 1983 年和 1984 年对价格影响程度的判断

资料来源：同表 1-1。

像例 1-1 中被调查人的性别的情况，属于双值分组。其他双值分组往往出现在以二分法（是否式问题）调查人们对一个问题的看法时。在研究问题时，有时需要将多值分组向双值分组的转换。例如例 1-1 调查对象对能源价格影响程度的判断本来是分为 4 组：“很严重”、“有些严重”、“不甚严重”和“根本不严重”。可以把前两组合并为“严重”，后两组合并为“不严重”，重新汇总数据得到表 1-5。这样可以更为简洁地展示 1984 年被调查者对能源价格影响严重性的判断强于 1983 年。