

顾问 王静龙 艾春荣 徐国祥 周 勇

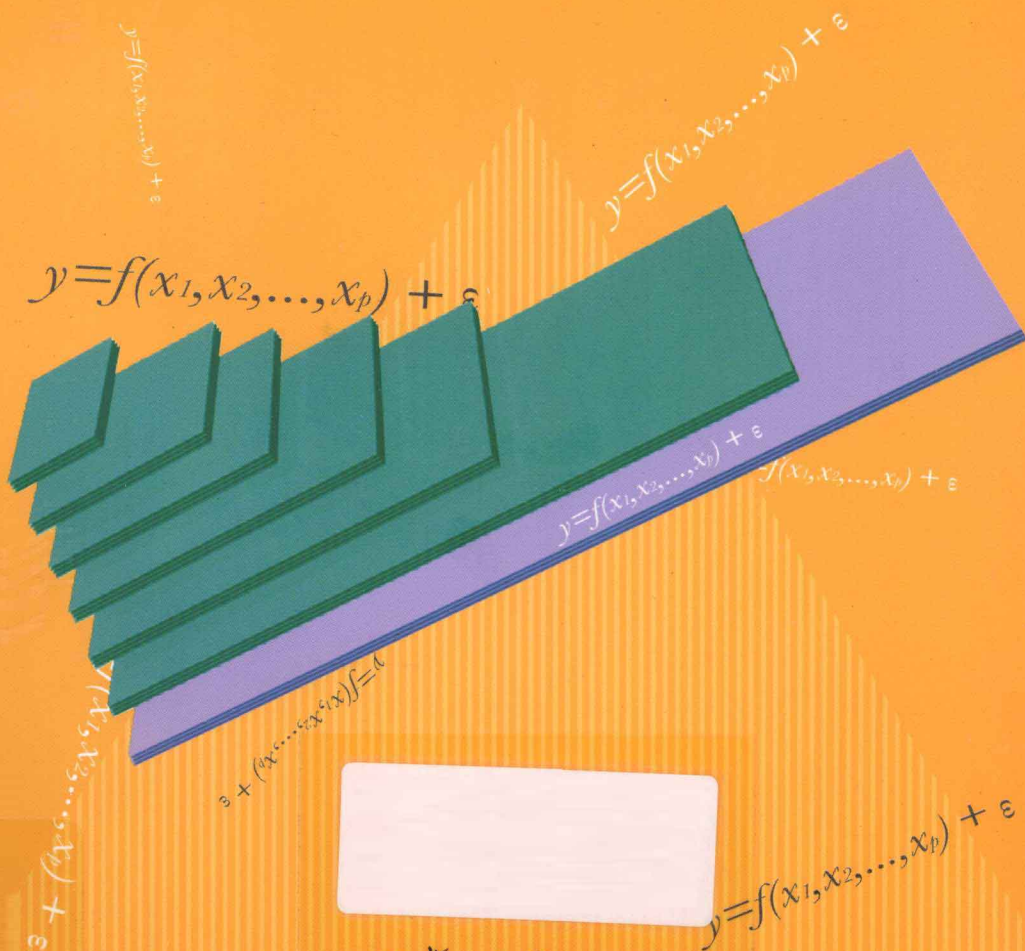
21世纪

高校统计学专业教材系列

数据、模型与决策

简明教程

王静龙 梁小筠 王黎明 编著



复旦大学出版社



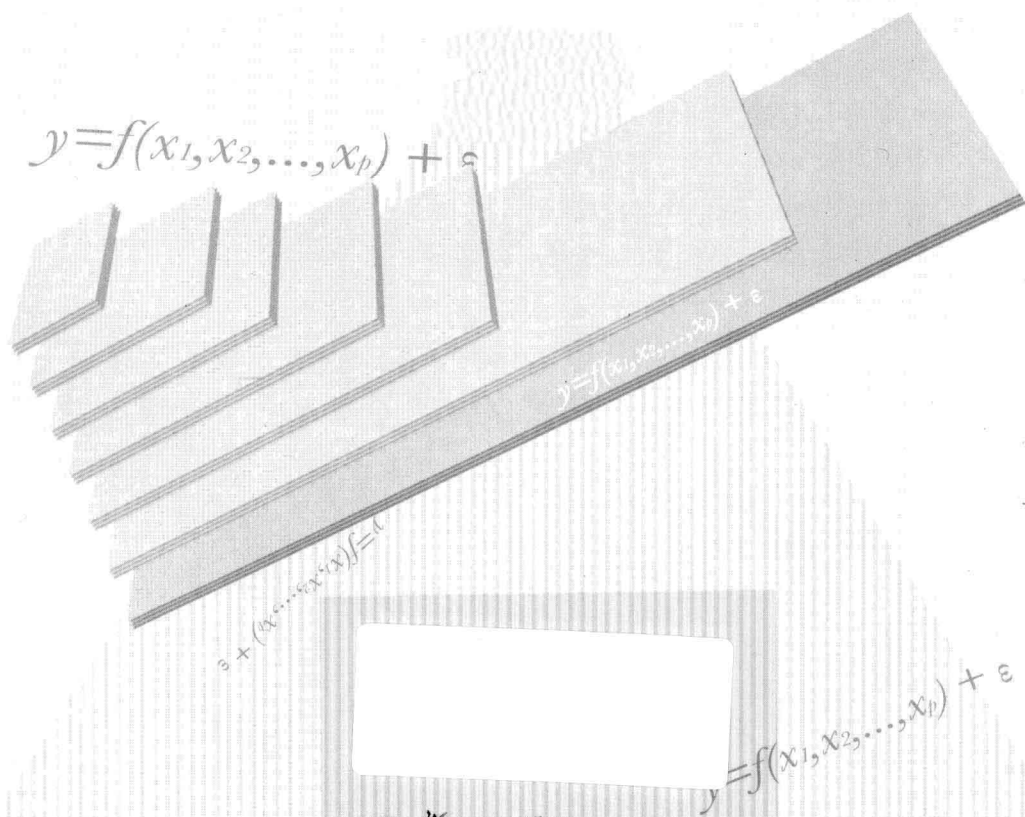
21世纪

高校统计学专业教材系列

数据、模型与决策

简明教程

王静龙 梁小筠 王黎明 编著



复旦大学出版社

图书在版编目(CIP)数据

数据、模型与决策简明教程/王静龙、梁小筠、王黎明编著. —上海:复旦大学出版社,2012.8
(复旦博学·21世纪高校统计学专业教材系列)
ISBN 978-7-309-09136-6

I. 数… II. ①王…②梁…③王… III. 决策模型-高等学校-教材 IV. C934

中国版本图书馆 CIP 数据核字(2012)第 174330 号

数据、模型与决策简明教程

王静龙 梁小筠 王黎明 编著
责任编辑/王联合

复旦大学出版社有限公司出版发行
上海市国权路 579 号 邮编:200433
网址:fupnet@fudanpress.com http://www.fudanpress.com
门市零售:86-21-65642857 团体订购:86-21-65118853
外埠邮购:86-21-65109143
上海浦东北联印刷厂

开本 787×960 1/16 印张 24.5 字数 443 千
2012 年 8 月第 1 版第 1 次印刷
印数 1—4 100

ISBN 978-7-309-09136-6/C·240
定价:45.00 元

如有印装质量问题,请向复旦大学出版社有限公司发行部调换。
版权所有 侵权必究

内容提要

这是一本专业教材。全书从统计、规划求解与决策分析三个模块，讲述从数据分析到模型建立，再到实践应用的过程。内容涵盖描述性统计分析，概率、随机变量及分布，估计问题，检验，多组数据比较分析，回归分析，规划求解，时间序列分析等。

这又是一本有趣的教材。为充分体现出数理统计在管理工作中的重要作用，本书不强调证明推导，讲究理论方法公式隐含着的思想。学员将案例分析、计算机数据实验与课堂内外交流讨论结合在一起，比如通过Excel，也能以极大的乐趣学会制表、画图、函数、数据分析与规划求解等知识。

本书可以作为经济学管理学的公共基础课教材，也非常适合MBA项目中的商务与经济统计课堂讲授。

本书向授课老师免费提供精致完整的PPT课件，用户可通过fudanjiacai@163.com索取。

总序

回顾新中国成立以来我国统计学科的发展道路,可以说是充满着曲折,也充满着希望。20世纪50年代照搬苏联模式,政府统计实务代替了统计学,否定了统计学作为一门方法论科学的存在。统计学的科学性和普遍应用性被曲解。这种情况直到1978年科学的春天来了才发生根本性的改变。党的改革开放政策使得统计学在中国得到了迅猛的发展。20世纪末教育部新颁的大学本科专业目录中,统计学被确立为一级学科。这就要求我们做好统计学专业课程建设的工作。为此上海财经大学统计学系与复旦大学出版社携手,会同国内相关院校及港台相关名校专家,策划出版一套主要针对统计学专业本科生使用的、适应新时期需要的系列教材——博学·21世纪高校统计学专业教材系列。

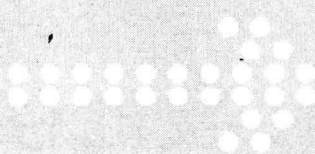
本教材系列力求体现以下特点:

第一,教材主要考虑面向财经类统计学专业,同时也要考虑“大统计学”专业的需求,力求选材做到“精”和“新”。

第二,内容选择将广泛吸收国内外优秀教材的成果,在系统介绍基本理论和基本方法的同时,注意介绍新的、成熟的内容,以及统计学在实际问题中的应用。

第三,教材编写注重计算机的应用,根据教材的具体内容选讲相应的统计软件,提高学生熟练运用统计方法和计算机技术解决实际问题的能力。

“博学·21世纪高校统计学专业教材系列”的前期规划教材包括《统计学》、《数理统计学》、《应用回归分析》、《国民经济核算原理》、《应用时间序列分析》、《统计计算》、《SAS数据分析系统教程》、《非参数统计学》、《变点统计分析及其应用》、《多元统计分析》、《金融计量统计学》;后期还将与境内外知名高校专家合作,陆续出版《抽样调查技术》、《贝叶斯统计学》、《实验设计与质量控制》、《统计预测与决策》、《高等数理统计学》和《金融时间序列分析》等。本教材系列的编写大纲和书稿



经过教材编写委员会的多次反复论证、认真讨论。感谢参与论证和编写的各位同行,希望他们的辛勤的劳动成果能够得到统计学界同行们的认可,获得同学的欢迎。这套系列教材的不当之处,恳请读者批评指正。为完善财经类统计学专业的教材建设,我们大家一起努力。

王静龙
2008年5月
于华东师范大学

前言

我第一次给 MBA 学员上课是 20 世纪 90 年代中期,在华东师范大学企业与经济发展研究所举办的 MBA 项目讲授商务与经济统计。自此之后,我陆续在各类 MBA 项目中,例如在华东师范大学商学院与荷兰马斯特里赫特管理学院合作举办的 MBA 项目中讲授商务与经济统计。21 世纪初起直到 2010 年我把在 MBA 教学过程中陆续编写的讲义整理修改,并参照全国工商管理硕士教育指导委员会审定的课程教学大纲,编写了“数据、模型与决策”这门 MBA 核心课程的讲义。2010 年下半年起我在我们学校的 MBA 项目使用这本讲义上课。在教学过程中充分听取学员的意见,对讲义进行了修改补充。本书就是在修改多次的讲义的基础上整理加工而成的。

学习“数据、模型与决策”这门课,每个 MBA 学员的基础不尽相同,差异之大出乎我的意料。学员中有的学理工,有的学文史,还有相当多的是学外语的,甚至有学艺术与体育的。不少学员没有学过高等数学。“数据、模型与决策”包括有统计、规划求解与决策分析 3 个模块。华东师范大学 MBA 核心课程的教学一共只有 10 次课(其中 1 次考试),每次 3 小时(包括休息)。课时数不多,学员基础差异不小,是这门课的一大困难。在我看来,完成这 3 个模块的教学,对我来说的确是个挑战,同时也给我带来乐趣。

一说学习“数据、模型与决策”这门课,学员们马上就想到数字、公式与计算。他们的反应就是,这门课肯定枯燥难学。有个学员说,我就是为了避开学习数学而选读外语专业的,想不到现在又要学数学。当然,MBA 学员学习这门课也有不少有利条件。虽然他们基础差异不小,但他们有管理的实践工作经验,知道量化分析在管理工作中的重要作用。学习这门课,在方法与公式的记忆能力方面他们不如在读大学生,但他们的理解能力不会比大学生差。

我在这门课的教学过程中,力求激发学员学习的兴趣,鼓起他们学习的信心,深刻认识这门课在管理工作中的广泛且重要的作用,体会量化分析的乐趣。授课



时注重讲清道理,不讲究证明推导,讲究的是理论方法公式隐含着的思想。学员学习注重应用,将案例分析、计算机数据实验与课堂内外的交流讨论结合在一起。计算机数据实验注重介绍学员熟悉的 Excel,学会操作它的制表、画图、函数、数据分析与规划求解等功能。前面的几次课难免需要讲述比较多的概念、术语与公式,因而学员说,听课开始犹如爬山,有点累。后面的课学员自己动手做的计算机数据实验越来越多,学员说我们在下山,轻松了。课程结束时不少学员说,体会到这门课的味道了,意犹未尽想再多学一些。

考虑到学员的可持续学习以及广大管理工作者的需要,本书的编写力求适合自学。为此尽量采用课堂讲授的语言与方式,叙述简要明确,重点突出。书中有很多例子。这些例子的解题思路尽量做到步骤清晰,便于读者领会模仿及开阔思路。除了第一章,每一章最后都有内容提要,点明本章学习内容与计算机数据实验的要求。每章配有与本章所学知识点有关的习题。本书的某一些章节标有星号“*”。如果教学时间紧可以跳过去,这并不影响全书的连贯。

本书第一与第二章由梁小筠教授执笔编写,第十章由王黎明教授执笔编写,其余各章由我编写。限于我们学识水平与 MBA 教学经验的不足,书中定有很多不足与不成熟的地方,恳请诸位同仁与 MBA 学员提出批评意见,以求不断改进。

感谢华东师范大学 MBA 中心的学员。如果没有这些学员的参与,本书难以成稿。感谢华东师范大学 MBA 中心的领导与老师的大力支持。最后,要感谢复旦大学出版社,没有他们的支持和高效率的工作,本书不可能很快出版。

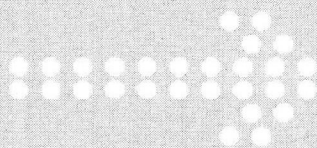
王静龙

2012 年 1 月 13 日

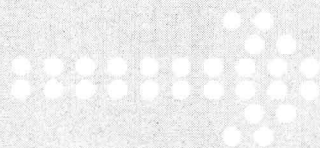
contents

目 录

第一章 引言	1
§ 1.1 数据	2
§ 1.1.1 调查数据	3
§ 1.1.2 观察数据	5
*§ 1.1.3 试验数据	10
§ 1.2 模型	17
习题一	20
第二章 描述性统计分析	25
§ 2.1 数据类别	25
§ 2.2 定性数据的描述性统计分析	27
§ 2.2.1 定性数据的列表描述	27
§ 2.2.2 定性数据的图示描述	31
§ 2.2.3 定性数据的数值描述	32
§ 2.3 定量数据的描述性统计分析	38
§ 2.3.1 定量数据的列表与图示描述	38
§ 2.3.2 定量数据的平均大小(集中趋势、中心位置)的数值描述	49
§ 2.3.3 定量数据的离散程度的数值描述	63
§ 2.3.4 经验法则	71
*§ 2.3.5 数据变换	75
内容提要	79
附 2.1 选择 Excel 的“数据”下拉菜单制表	80
附 2.2 选择 Excel 的工具下拉菜单,使用数据分析功能制表和画图	81
附 2.3 统计量与 Excel 函数	82
附 2.4 选择 Excel 的工具下拉菜单,使用数据分析功能计算描述统计量	82



习题二	83
第三章 概率、随机变量及其分布	88
§ 3.1 概率	88
§ 3.1.1 古典概率:利用对称性计算概率	89
§ 3.1.2 频率方法:估算概率	92
§ 3.1.3 主观推测:估算概率	96
§ 3.2 随机变量及其分布	99
§ 3.2.1 离散型随机变量	99
§ 3.2.2 离散型随机变量的平均数(均值、期望)	102
§ 3.2.3 离散型随机变量的方差与标准差	105
§ 3.2.4 独立性	110
§ 3.3 期望-方差的决策分析	110
§ 3.4 常用的离散型随机变量的分布	113
§ 3.4.1 二项分布	113
*§ 3.4.2 泊松分布	122
§ 3.5 正态分布	125
§ 3.5.1 正态曲线	127
§ 3.5.2 正态分布均值、方差与标准差的估计	130
§ 3.5.3 正态分布的 Excel 函数命令	133
§ 3.5.4 标准正态分布	135
*§ 3.5.5 正态分布很重要	138
内容提要	140
附 3.1 二项分布、泊松分布与正态分布的 Excel 函数命令	142
习题三	143
第四章 估计问题	149
§ 4.1 比例的估计	149
§ 4.1.1 抽样调查的可信性	151



§ 4.1.2 估计比例	153
§ 4.1.3 支持度估计大的候选人是否支持度也大	159
*§ 4.1.4 比例之差的区间估计	162
§ 4.2 均值的估计	164
§ 4.2.1 大样本情况	165
§ 4.2.2 小样本情况	166
内容提要	171
习题四	173

第五章 检验	176
§ 5.1 统计检验问题的推断思考过程	176
§ 5.2 统计检验问题的求解	180
§ 5.2.1 原假设与备择假设	180
§ 5.2.2 检验法则与 p 值	181
§ 5.2.3 检验的水平	183
*§ 5.2.4 两类错误	185
§ 5.2.5 假设检验问题的求解步骤	186
§ 5.3 比例的检验	187
§ 5.4 正态分布均值的 t 检验	190
*§ 5.5 正态分布方差的 χ^2 检验	196
*§ 5.6 属性数据类别比例的 χ^2 检验	199
§ 5.6.1 属性数据类别比例的检验问题	200
§ 5.6.2 齐性检验	203
内容提要	208
习题五	209

第六章 多组数据的比较分析	212
§ 6.1 两组数据的比较	212
§ 6.1.1 比较正态分布方差的 F 检验	213

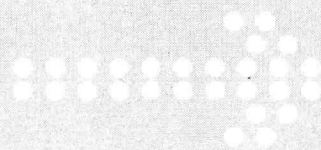
§ 6.1.2 比较正态分布均值的 t 检验	217
§ 6.1.3 成对数据	226
§ 6.2 方差分析(ANOVA)	228
*§ 6.2.1 组间变差	230
*§ 6.2.2 组内变差与全变差	232
*§ 6.2.3 变异分解	235
*§ 6.2.4 多重比较	236
内容提要	237
附 6.1 用 Excel 的数据分析功能比较两组数据	239
附 6.2 用 Excel 的数据分析功能进行方差分析(ANOVA)	239
习题六	240

第七章 相关与回归分析	244
§ 7.1 相关与回归	244
§ 7.1.1 正相关关系与正比例关系	244
§ 7.1.2 负相关关系	245
§ 7.1.3 儿子身高的回归	245
§ 7.1.4 预测孩子成年后身高	248
§ 7.1.5 回归模型	249
§ 7.2 建立回归模型	250
§ 7.2.1 收集数据	251
§ 7.2.2 判断变量之间有没有相关性	251
§ 7.2.3 计算回归直线	255
§ 7.2.4 给出预测误差和概率	256
§ 7.3 相关系数	260
§ 7.3.1 检验:是否线性相关	262
§ 7.3.2 测定系数	264
*§ 7.3.3 相关系数的计算公式的由来	264
§ 7.4 多元线性模型	266

§ 7.4.1	回归模型复相关系数、测定系数与回归标准误	267
§ 7.4.2	检验回归模型有没有意义	269
§ 7.4.3	截距与斜率是否等于 0 的检验以及它们的估计问题	271
*§ 7.4.4	将非线性回归问题化为线性回归问题	273
*§ 7.4.5	复共线性	278
内容提要		282
习题七		284

第八章 规划求解		289
§ 8.1 线性规划		289
§ 8.1.1 最大利润规划问题		290
§ 8.1.2 线性规划问题的解		291
§ 8.1.3 敏感性分析		296
§ 8.1.4 最小运输成本规划问题		299
§ 8.2 非线性规划		303
§ 8.2.1 投资组合的回报		304
§ 8.2.2 优化投资组合		305
§ 8.3 整数规划		311
内容提要		312
习题八		313

第九章 决策分析		316
§ 9.1 不确定型决策问题常用的决策准则		317
§ 9.2 风险型决策问题常用的决策准则		320
§ 9.2.1 期望收益决策准则		320
§ 9.2.2 期望后悔(机会损失)决策准则		321
§ 9.2.3 最大可能决策准则		322
§ 9.3 贝叶斯决策		322
§ 9.4 效用函数		326



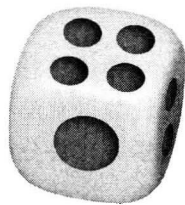
§ 9.5 德尔菲法与层次分析法	329
§ 9.5.1 德尔菲法	329
§ 9.5.2 层次分析法	330
§ 9.6 博弈论	337
§ 9.6.1 纳什(Nash)均衡	338
§ 9.6.2 混合策略	341
内容提要	343
习题九	343
第十章 时间序列分析	345
§ 10.1 简单时间序列分析	345
§ 10.1.1 时间序列趋势外推法	345
§ 10.1.2 趋势预测模型的确定	346
§ 10.1.3 移动平均	353
§ 10.1.4 指数平滑	356
*§ 10.2 平稳时间序列模型	358
§ 10.2.1 自相关函数(AFC)和偏自相关函数(PAFC)	359
§ 10.2.2 自回归模型(Autoregressive Model)	364
§ 10.2.3 移动平均模型(Moving Average Model)	366
§ 10.2.4 自回归移动平均模型(Autoregressive-Moving Average Model)	367
§ 10.2.5 平稳时间序列建模	368
§ 10.2.6 平稳时间序列预测	373
内容提要	374
习题十	375
参考文献	377

第一章

引 言

各行各业的管理人员经常需要在不确定的情况下做出决策。不确定性情况又称不确定性现象,就是不是只有一个结果的现象。例如,某公司正在考虑要不要开发一个新产品。倘若决定开发新产品,则工厂需要扩建。扩建工厂有两种选择:中型扩建与大型扩建。开发还是不开发新产品? 中型扩建还是大型扩建为好? 公司管理人员难以作决策的一个很重要的原因是,市场对该新产品需求量究竟如何? 市场对该新产品的需求往往是不确定的,例如有高、中与低 3 种状态。由此可见,市场需求是不确定性现象,它有 3 个不同的状态:高、中与低。

不确定性现象,又称**随机现象**。随机现象也有规律性,例如根据我们的观察,随机现象的这一个结果与另一个结果发生的可能性是一样大,也可能有大小之别。不确定性现象某个结果发生的可能性的数量大小可以用一个数来度量。我们以人们熟知的骰子(见右图)为例,说明随机现象的数量规律性。抛掷一颗骰子,观察骰子落地时正面出现的点数。通常称这样的试验为**随机试验**。



这个随机试验有 6 个可能出现的结果:出现 1 点、2 点、3 点、4 点、5 点和 6 点。众所周知,对于一颗均匀的骰子来说,出现这 6 种情况的可能性是一样大的,都等于 $1/6$ 。当然,在实际管理业务工作中,如骰子那样,很容易判断出可能性的情况是很稀罕的。例如,市场需求高、中和低的可能性的推测与判断,众人的看法不尽相同。即使两人都主管销售,但由于他们有各自的经验,看法也可能有差别。倘若对市场需求的可能性的推测与判断经过讨论取得了一致的意见,那么该如何决策? 要不要开发新产品? 是中型扩建还是大型扩建? 显然,管理决策问题的妥善解决,需要各种类型的定性分析与定量分析方法的交织应用。本课程研究定量分析,研究不确定性现象的数量规律性。

从数据出发,建立适当的模型,对所考虑的问题作出推断,为管理者的决策提供依据和建议。

本课程将讲解统计学、规划论与决策论的基本理论和方法,以及它们是如何用于定量分析的。本课程的讲解尤以统计学的基本理论和方法为主。早期,统计

主要是为统治国家服务的。英文 statistics(统计)是由拉丁文 status(状态、国家)和 statista(政治家)演化而来。近代随着现代化大生产的发展,统计应用的范围越来越广,扩展到各行各业。在我国,尤其是改革开放以来,统计活动空前活跃,统计的重要性越来越被大家重视。这正如我国著名学者马寅初先生(1882—1982)所说的:

学者不能离开统计而研究,
政治家不能离开统计而施政,
企业家不能离开统计而执业。

§ 1.1 数据

英国的《不列颠百科全书》对统计给出了一个简单明了且形象生动的定义:

统计是收集和分析数据的科学与艺术。

这句话告诉我们统计学的研究对象是数据。统计能做两件事:一是收集数据,二是分析数据。俗话说,巧媳妇难做无米之炊。那么,对统计来说

数据! 数据! 数据! 我们不能做无米之炊。

必须注意的是,巧媳妇可以等丈夫买来米之后下锅,我们却不能等米下锅,不能说数据有没有不是我们统计的事,因而认为所谓统计工作就是有了数据之后用数学方法去分析数据,这样一种认识是误解了统计。分析数据仅仅是统计的一项工作。统计还有另一项工作,那就是收集数据。而且收集数据所花的时间、财力与精力不比分析数据少,甚至更多。下面这样的一种说法并不过分,收集到质量高(能说明问题)的数据比分析数据更为重要。为什么这样说,那是因为**数据是决策的一个非常重要的依据**。下面的两个例子说明数据的重要性。

商场选址

上海克莉丝汀食品店选址,与地图方位相比,他们更看中新店铺实际可吸纳的客流量,有专人用计表器测算可能的客源,达到一定的数量后公司网点开发部才会提交报告。这是他们网点选址前要做的必要的定量分析。除计表器测客流量外,克莉丝汀有一整套的扩张全攻略。他们做网点规划,首先根据该城市居民人均可支配收入考虑网点数;根据公司的定位,网点在客流量最集聚的地方,例如南京路、淮海路和豫园商场并不多见,网点主要是跟着社区走,沿着地铁走,而且很多店铺都选在十字路口拐角处,四面都可以看到克莉丝汀的招牌。

肯德基在开一家新店前会进行如下的商圈研究:

- 周边情况:周边商厦、饮食店、学校、娱乐场所、公交站点的数量和分布等;

- 半径范围:从拟建的餐厅出发,朝不同方向步行半分钟、1 分钟、1 分半钟乃至 5 分钟路程内到达上述这些设施的准确位置;
- 人流量测试:在拟建的餐厅处,周一至周日,每天按不同时段严格统计人流量。

产品的认知度与购买意愿

康泰克“早一粒,晚一粒”的广告用语曾经家喻户晓。但自 2000 年 11 月 15 日起,由于含有 PPA(苯丙醇胺)经媒体曝光,占国内感冒市场六成的康泰克突然退出。这给天津中美史克制药有限公司带来了巨大的损失。但仅事隔 9 个月,2001 年 8 月,不含 PPA,代之以 PSE(盐酸伪麻黄钙),同时保留扑尔敏成分的新康泰克上市。中美史克敢于耗资 1.45 亿元上马新康泰克,欲重登国内感冒药市场冠军宝座,这和它们前期所做的市场调查有关。北京美兰德信息公司对北京、上海等 20 座城市的感冒药市场调查表明,康泰克在全国享有 89.6% 的认知度,90% 的被调查者表示“会接受”或“可以接受”康泰克重回市场。正因为确信 90% 的人知道康泰克,确信 90% 的人愿意在不含 PPA 之后继续使用康泰克,所以中美史克确信康泰克可以复活。

企业在开发新产品前都要进行市场调查,了解顾客对品牌的认知度与购买意愿等问题。所谓品牌的认知度就是对品牌内涵及价值的认识与理解程度。市场调查与研究的内容非常丰富,例如,调查什么问题,向谁调查,调查多少人,如何保证调查质量,以及调查数据的处理分析等,这些都要用到统计的理论和方法。

我们这里所说的数据并不仅仅是数值的意思。数据的英文名是 data,它是拉丁文 datum 的复数形式,其含义简单地说是“事实资料”。本课程所说的

数据既包括数值型资料,也有文字型资料。

前面我们说,英国的《不列颠百科全书》给出了统计一个简单明了且形象生动的定义:统计是收集和分析数据的科学与艺术。下面简要介绍收集数据的调查、观察和试验这 3 个方法,以及在收集数据时应该注意的一些问题。

§ 1.1.1 调查数据

调查有两种类型:普查(又称全面调查)与抽样调查。一般来说,普查是为某个特定目的专门组织的某个地区、甚至全国的一次全面调查。我国在 1953 年组织了第一次全国人口普查,接着于 1964 年与 1982 年分别组织了第二与第三次全国人口普查。在第三次全国人口普查之后我国按世界各国的惯例,逢“0”年进行人口普查。1990 年、2000 年与 2010 年分别组织了第四、第五与第六次全国人口