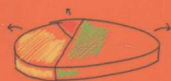


看穿一切数字的 统计学

[日] 西内启◎著 朱悦玮◎译



統計学が最強
の学問である



中信出版社 CHINA CITIC PRESS



統計学が最強
の学問である

統計学

看穿一切数字的

〔日〕西内启◎著 朱悦玮◎译



NLIC2970935664

图书在版编目 (CIP) 数据

看穿一切数字的统计学 / (日) 西内启著; 朱悦玮译. —北京: 中信出版社, 2013.10

书名原文: 統計学が最強の学問である

ISBN 978-7-5086-4206-2

I. ①看… II. ①西… ②朱… III. ①经济统计学-通俗读物 IV. ①F222-49

中国版本图书馆CIP数据核字 (2013) 第 208766 号

TOKEIGAKU GA SAIKYO NO GAKUMON DE ARU

Copyright © 2013 Hiromu Nishiuchi

Simplified Chinese translation copyright ©2013 by CHINA CITIC PRESS.

All rights reserved.

Original Japanese language edition published by Diamond, Inc.

Simplified Chinese translation rights arranged with Diamond, Inc.

through Beijing GW Culture Communications Co., Ltd.

本书仅限中国大陆地区发行销售

看穿一切数字的统计学

著 者: [日] 西内启

译 者: 朱悦玮

策划推广: 中信出版社 (China CITIC Press)

出版发行: 中信出版集团股份有限公司

(北京市朝阳区惠新东街甲 4 号富盛大厦 2 座 邮编 100029)

(CITIC Publishing Group)

承 印 者: 北京诚信伟业印刷有限公司

开 本: 880mm×1230mm 1/32

版 次: 2013 年 10 月第 1 版

京权图字: 01-2013-6209

书 号: ISBN 978-7-5086-4206-2/F · 2997

定 价: 36.00 元

印 张: 9.75 字 数: 131 千字

印 次: 2013 年 10 月第 1 次印刷

广告经营许可证: 京朝工商广字第 8087 号

版权所有 · 侵权必究

凡购本社图书, 如有缺页、倒页、脱页, 由发行公司负责退换。

服务热线: 010-84849555 服务传真: 010-84849000

投稿邮箱: author@citicpub.com

目 录
統計学が最強
の学問である

第一章

为什么统计学是最牛的学问？

- 01 没有统计学思维的人将会吃大亏的时代 / / 003
- 02 统计学是能够得出最好、最快答案的正确方法 / / 009
- 03 所有的学问都从统计学开始 / / 017
- 04 大数据时代，统计学火了 / / 025

第二章

大幅减少信息成本的抽样调查

- 05 统计学家的大数据“狂想曲” / / 037
- 06 准确度令人震惊的抽样调查 / / 043
- 07 为了 1% 的准确率值得花费几千万吗？ / / 051

第三章

统计学的关键：误差与因果关系

- 08 抛弃传统的统计方法吧！ / / 061
- 09 单纯收集数据会得出“禁食米饭”的荒谬结论 / / 067

- 10 改变投放广告单的方式就能赚取 60 亿日元 / / 073
- 11 真的增加了一亿销售额，还是只是误差？ / / 081
- 12 值得分析的数据都是与利益明显相关的 / / 089
- 13 统计学中的重大问题：“因果关系的方向” / / 095

第四章

最强的统计学武器：随机对照试验

- 14 先放牛奶还是先放红茶？ / / 103
- 15 随机对照试验：商业竞争中的有力武器 / / 113
- 16 “一次购买两台缝纫机可以打 9 折”能够增加销量吗？ / / 121
- 17 随机化面对的现实、伦理和感情之壁 / / 129

第五章

无法进行随机对照试验时该怎么办？

- 18 低成本、快速搜集数据的流行病学方法 / / 139
- 19 “回归平凡”的回归分析 / / 149
- 20 所有领域都可以应用的回归分析法 / / 159
- 21 快速加深对统计学的理解的一张图表 / / 171

- 22 现代统计学的主角：多元回归分析 / / 181
- 23 统计学家最拿手的统计法 / / 195

第六章

应对一切问题的统计学思考方法

- 24 统计学家们的残酷战争 / / 207
- 25 发明“IQ”的心理统计学 / / 217
- 26 诞生于市场调查现场的数据挖掘 / / 231
- 27 分析语言的文本挖掘：莎士比亚真的存在吗？ / / 247
- 28 “演绎”的计量经济学与“归纳”的统计学 / / 257
- 29 贝叶斯派与频率派之间的对立 / / 269

第七章

帮你站在巨人肩膀上的统计学

- 30 用统计学寻找“最佳答案” / / 283
- 31 寻找实证来结束争论 / / 293

后记 / / 303

- 第一章

为什么统计学是最牛的学问？

- 統計学が最強の学問である

● 統計学が最強
の学問である

01

没有统计学思维的人将会吃大亏的时代

赫伯特·乔治·威尔斯关于统计学的预言

“1903 年，赫伯特·乔治·威尔斯曾经预言，在未来社会，统计学思维将像阅读能力一样成为社会人必不可少的能力。”哈佛大学医学院的统计学教材的开篇写着这样一句话。

赫伯特·乔治·威尔斯是被称为“科幻小说之父”的作家和思想家。在威尔斯的著作中首次出现了时间机器和透明人等科幻概念，他凭借自己丰富的科学与预见性，准确地预言了核武器与联合国乃至现在被我们称为维基百科的百科事典的出现。

在现代统计学还处于黎明期的 1903 年，威尔斯为何做出这样的预言，我们无从得知。但是在接近 100 年后的今天，统计学的

思考方法对我们来说毫无疑问已经成为与阅读能力同样重要的能力。就好像一个没有阅读能力的人在现代社会寸步难行，没有统计学思维的人同样难以在现代社会生存。

没有阅读能力就无法理解合同和法律中的内容，没有统计学思维就无法了解概率和数据。这两者缺一不可，否则就会陷入毫无防备的状态，成为那些合法诈骗的受害者，而你自己又无法抱怨。

让抽签必胜的统计学思维

举一个我身边的例子，在上大学的时候，我经常与研究室的朋友抽签决定谁去便利店买东西。

抽签有很多种方法，那天我们选择的方法是由4名参与者画出8条竖线，然后我在其他人都看不见的情况下，从左边数起，在第4条竖线下面画了一颗星星，剩下的3人则在我看不见的情况下，每人画4条横线。参加者猜拳决定顺序，从1~8条竖线中选择一条，一轮结束后顺序调转过来继续。选中星星符号的人要替其他人去便利店买东西（图1-1）。

要是全凭直觉来玩这个游戏，那么你获胜的概率一定很低。

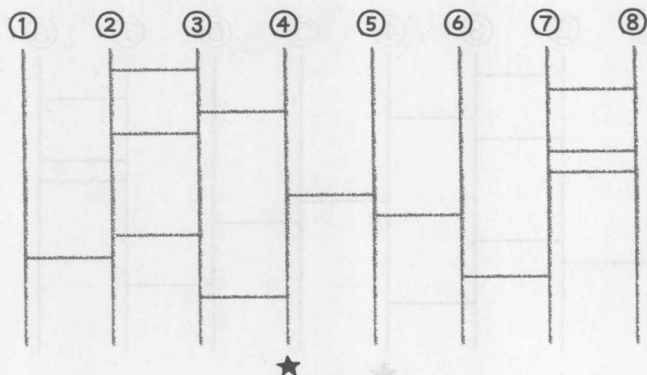


图 1-1

我们尝试按照这个规则重复 1 000 次，将每条竖线选中星星的次数进行模拟试验，结果如图 1-2 所示。

选中星星概率最高的是正上方的 4，在 1 000 次的试验中选中了 210 次，也就是说有 21.0% 的概率。紧接着，就是右边的 19.4% 概率。最右边的概率最低，只有 3.3%。

实际上，在类似这样的抽签游戏中，或许是出于人类的心理倾向，最先选择两端竖线的人少之又少。

也就是说，看上去好像是每人 $1/4$ ，也就是 25% 概率的公平比赛，但实际上一直选择两端的我可能去便利店的概率只有 11.4% [= (81+33) / 1 000]。凭直觉选择靠近中心竖线的朋友，

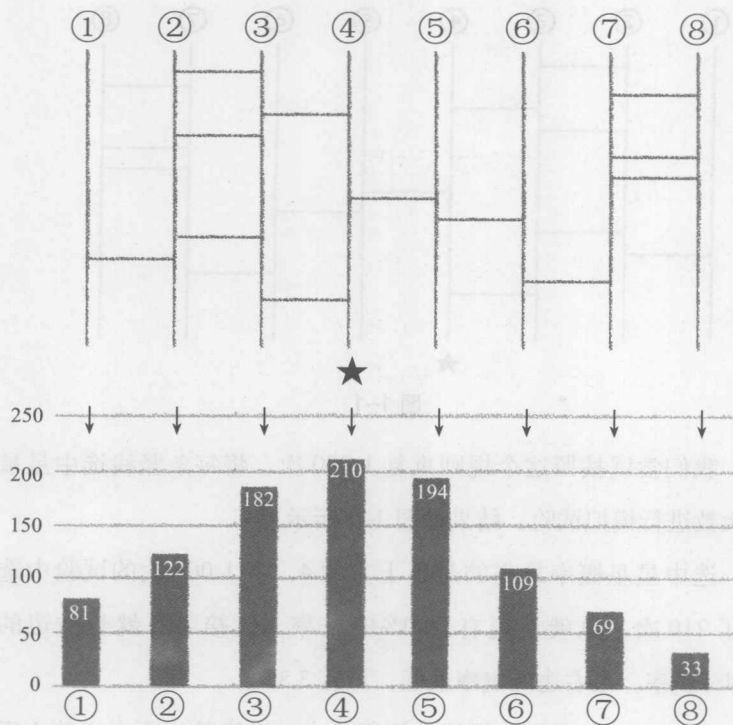


图 1-2

则有 40.4% [= (210+194) / 1000] 的概率选中。对于抽签的结果，他们每次都会带着“最近运气真差”的感觉去便利店买东西。

另外，似乎有的地方政府在对公共工程进行招标时的最后选

择是“相同条件的情况下采用抽签方式作决定”，也许有的公司就利用概率知识提高了公司的中标率。

掌握统计学的人就能够掌握世界

当然，画横线的过程就像模拟试验一样是完全随机的，所以不管概率有多低，我也不能完全规避。但是，只要掌握一定程度的统计学，就能够在这种不确定的情况下稍微耍一点手段，抢得先机。

比如，我为一家零售企业作数据分析，我将他们之前漫无目的投放直接邮寄广告（DM）的目标群体，分成了“应该发送的客户”和“不应该发送的客户”，经过优化选择后他们的销售额增加了6%。因为他们之前的销售额是1 000亿日元，增加的6%也就是60亿日元。

我并没有增加DM的投放量，所以没有增加成本，只是通过分析找出发送DM后“能够增加销售额的顾客”和“不会增加销售额的顾客”，就好像在抽签中所使用的手段一样，使销售额提高6%的“手段”也是存在的。

假如你和你的公司没有使用这种手段，而竞争对手的公司却

使用了，那就会抢走你的顾客与利益。

对于生存在 21 世纪的我们来说，统计学已经是必不可少的技能，同时对于很多人来说也是最强的武器。在商业领域应用统计学解决问题被称为商务智能，其中的智能指的就是经常出现在谍战电影中的中央情报局（CIA）的“I”。而且，从中国古代的《孙子兵法》开始，就一直强调战争中情报的重要性。

掌握情报的人就能够掌握世界，这句话放在现代社会，就应该是“掌握统计学的人，就能够掌握世界”。

02

统计学是能够得出最好、最快答案的正确方法

统计学成为最强武器的原因

为什么统计学会成为最强的武器呢？

用一句话来回答，那就是不管在什么领域，收集数据进行分析都是得出最好、最快答案的方法。

如果你在一家零售企业工作，总裁打算召开全体会议研究如何增加销量，那么这究竟会是怎样一种情况呢？

首先，先对增加销量提出方案，然后那些老资历的员工就会主观地提出“我感觉……”或者“根据我常年的经验……”之类的建议。

我可以断言，如果你的公司拥有足够多的数据，但是却不

对数据进行分析，而只是根据直觉和经验来进行讨论的话，那完全是在浪费时间。不仅如此，还会白白地浪费会议参加者的人工费。日本有很多公司，对于那些时薪为 800 日元的年轻人非常苛刻，经常呵斥他们偷懒，但是却对那些时薪高于他们几倍甚至几十倍的老员工在会议上浪费时间的情况毫不在意。

单凭感觉做出的决定导致犯错误的情况，会给经营者、员工以及客户造成困扰，但是在这个世界上有些情况是绝对不允许出现错误的，比如说人命相关的情况。

如果你的一个错误选择会导致 10 万人失去性命的话，你究竟应该以什么为依据做出判断呢？在这种时候，你还会依靠领导的直觉和经验来做出决定吗？

或者说，像日本首相或者日本厚生劳动大臣这些手握重权的人，如果由他来决定包括你和你的家人在内的 10 万人的性命，他若是只凭借毫无根据的个人意见来进行决定的话，你会是什么感觉呢？

一个错误的判断就会导致 10 万人失去生命，这种状况像是科幻电影中出现的场面。但是在医疗领域，特别是我从事的“公共卫生”以及“社会医学”、“保健行政”等领域，这类情况却时有发生。

日本每年有 35 万人死亡，其中约 19 万人死于心脏病，约 3 万人死于自杀。但是，如果能够采取适当的预防和治疗措施，就可以拯救其中几万人的性命。

为了在面对像这样关系到大量人命的问题时不会做出错误的选择，在 19 世纪的伦敦，人类历史上第一次利用统计学的力量与夺走数以万计人类性命的原因展开了战斗。

为了遏止原因不明的流行病而诞生的学问叫作“流行病学”，世界上最早的流行病学研究开始于 19 世纪的伦敦，针对的是一种叫作霍乱的流行病。在流行病学之中，统计学也发挥了巨大的作用。

当时霍乱在英国本土爆发了 4 次，据说总共夺走了十几万人的性命。尽管当时的科学技术没有现代这么先进，但是当时的伦敦也聚集了大量接受过高等教育的科学家和医生，还有优秀的政府官员。虽然他们都是非常聪明且能力很强的人，面对霍乱的流行却一筹莫展。甚至可以说，他们的一些决定对当时的局面是有害无利的。

比如，有的医生提议用他调制的特别除臭剂来阻止霍乱的流行。

当时的伦敦正处于工业革命的高潮期，在这一时期，依靠农业无法养家糊口的人全都挤进城市，进入工厂变成劳动者。城市