

数据的真相

如何在数字
时代做出
明智决策

EVERYDATA

The Misinformation Hidden
in the Little Data
You Consume Every Day

MIT科学家
解读数据的本质

消费者及市场、营销、
广告、公关从业人员相见恨晚的一本书

[美] 约翰·W·约翰逊
[美] 迈克·格鲁克 著

王喆 译

营销大师

赛斯·高汀
推荐

海外借

中信出版集团

数据 的 真相

如何在数字时代做出
明智决策

[美] 约翰·H.约翰逊 [美] 迈克·格鲁克◎著

王喆◎译



图书在版编目 (CIP) 数据

数据的真相：如何在数字时代做出明智决策 / (美)

约翰·H. 约翰逊, (美) 迈克·格鲁克著; 王喆译. --

北京: 中信出版社, 2018.7

书名原文: EVERYDATA: The Misinformation Hidden

in the Little Data You Consume Every Day

ISBN 978-7-5086-8998-2

I. ①数… II. ①约… ②迈… ③王… III. ①数字技
术-应用-市场营销-研究 IV. ①F713.50

中国版本图书馆CIP数据核字 (2018) 第 110238 号

EVERYDATA: The Misinformation Hidden in the Little Data

You Consume Every Day by John H. Johnson, PhD and Mike Gluck

Copyright © 2016 by John H. Johnson, PhD and Mike Gluck

First published by Bibliomotion, Inc., Brookline, Massachusetts, USA.

This translation is published by arrangement with Bibliomotion, Inc.

Simplified Chinese translation copyright © 2018 by CITIC Press Corporation

ALL RIGHTS RESERVED

本书仅限中国大陆地区发行销售

数据的真相——如何在数字时代做出明智决策

著 者: [美]约翰·H. 约翰逊 [美]迈克·格鲁克

译 者: 王喆

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)

承 印 者: 北京诚信伟业印刷有限公司

开 本: 880mm × 1230mm 1/32

印 张: 9 字 数: 170 千字

版 次: 2018 年 7 月第 1 版

印 次: 2018 年 7 月第 1 次印刷

京权图字: 01-2018-2594

广告经营许可证: 京朝工商广字第 8087 号

书 号: ISBN 978-7-5086-8998-2

定 价: 49.00 元

版权所有·侵权必究

如有印刷、装订问题, 本公司负责调换。

服务热线: 400-600-8099

投稿邮箱: author@citicpub.com

2010年春，美国国家橄榄球联盟（NFL）的老板们在提出诸多要求之外，还想将常规赛季的比赛场数额外增加两场。作为国家橄榄球球员协会的执行主任，我觉得这并不是一件好事。橄榄球赛的受伤率已经高达100%，因此多打比赛意味着更多伤病。而伤病率的上升会导致运动员职业生涯进一步缩短，现在运动员的运动生涯本就只有三年多一点。橄榄球运动员在赛季期间每周都会冒着很大风险投身于这项美国人民最为喜爱的运动中，我们协会对他们的安全、报酬以及保护措施极为关注。

职业橄榄球产业创造的价值高达数十亿美元，而且依然在不断发展壮大之中。但在2010—2011年间，该行业经历了数十年不遇的停摆，因此我需要以一个具有说服力的方式来证明赛季的长度必须维持在十六场比赛不变。

于是我拨通了约翰·约翰逊的电话。

约翰是经济学、统计学、数据学领域的专家，他能够把极为复杂的数据理论变得简单易懂，让人一目了然。简而言之，他应该能够算得上是我认识的人中最为聪明的一个。他能够仔细、全面地解释问题，不论你受教育水平如何，都能够听得懂。而且最重要一点，他这人还很风趣。

怀着和美国所有橄榄球迷一样的心愿，国家橄榄球球员协会想拯救即将到来的全国橄榄球联赛，因此我们把相关数据交给了约翰。约翰及其同事开发出一个模型，展示球员受伤的时间和频率。他们指出导致最严重伤害的动作（从而引出对球员安全造成巨大影响的“开球规则”）。他们通过数据预测出平均每个球员职业生涯缩短的程度，估计出将全国橄榄球联赛赛季延长至十八场比赛可能导致的经济损失。他们同我们合作，将每场全国橄榄球联赛对周边社区产生的实际价值进行量化，帮助我们对可能发生的结束赛季停摆所产生的经济影响有一个更好的理解。

最终，约翰以其出色的口才向球员、媒体，尤其是美国国家橄榄球联盟的老板们（他们对“损失”这类词特别关心）解释了这些数据，我们得以做出既符合我们初衷，又能保障球员的健康和安全的决定，最终促成了后续的谈判，签订了一项历史性的协议，该协议为期十年，很好地保障了美国国家橄榄球联盟球员的权益。

本书很好地反映了约翰的思想。我对该书能够帮助你在日

常生活中做出更好的决定十分有信心。通过阅读本书，你可以学到如何对每天接收到的数据进行理解、解读、思考。约翰和本书的合著者迈克·格鲁克通过列举数百个例子，直指问题本质，举重若轻地将复杂问题化繁为简（有时候甚至能让你忍俊不禁）。因此在阅读本书过程中，要是发现自己因为恍然大悟而频频点头，并开始质疑家里和单位中看到的每一项“事实”，请不要感到惊讶。

有一个不可否认的事实，即你每天生活中的数据正在迅速增加，并从四面八方向你涌来。因此了解如何利用数据，以及在什么情况下数据会被滥用十分重要。我亲身体验了数据的力量，但你不需要和一群世界上最富有的人一起死命加班，体会理解数据、管理数据的重要性。在你看电视、购物、工作、在餐厅吃饭的时候，这本书是你的不二之选。《数据的真相》见解深刻、文字富有魅力、内容引人入胜，在关键时刻填补了一项关键空白。

本书相当于橄榄球队主教练编写的比赛战略手册，能够助你在运用数据的赛场上变得更强大、更聪明、更有信心。

祝阅读愉快。

莫里斯·F. 史密斯

美国国家橄榄球球员协会
执行主任

一万个问题亟待解决

Everydata
序
The Misinformation
Hidden in the Little
Data You Consume
Every Day

见鬼，明尼苏达州究竟发生了什么事？

2015 年夏天，10 天内有关方面宣布召回三种被沙门氏菌污染的食品。这三类问题食品的源头都在明尼苏达州，当地人吃坏了肚子。

涉事企业最终召回了约 370 万磅鸡肉制品，除此之外还有数量不明的黄鳍金枪鱼¹。

但是，为什么是明尼苏达州？

是不是所有受到污染的食物都产自明尼苏达州？不是。那是不是都从那儿装船运输？不是。是不是明尼苏达州的居民更容易感染沙门氏菌导致的疾病？据我们所知，也并非如此。

据雅虎健康网的报道，此事件的内部联系十分简单：明尼苏达州和美国其他州相比，在诊断由食物引起的疾病方面水平更高。²

有些人可能听说有三起食品召回事件都发生

在明尼苏达州，就认为那个地区的食品不安全。但事实并非如此。恰恰相反，由于明尼苏达州健康农业部工作出色，你的健康在那里可能比在美国其他地区得到更好的保障。

每天你都被媒体报道或其他来源的消息所包围，这些消息通常充满了隐藏信息——或误导信息。本书将帮助你辨别、解读这些信息，使你成熟地汲取数据（“信息”或“事实”比较高端的说法）。

全书通篇将回答以下问题：

一则虚假新闻如何让股市一下子蒸发掉 1 360 亿美元？

（第六章）

福岛核灾难是否可以避免？（第八章）

4/5 的儿科医生是如何看待婴儿食品的？（第七章）

如何知道哪个总统候选人的选票真正领先？（第五章）

吃烤奶酪三明治的人们真的更为“性福”吗？（第四章）

“小数据”不是一个常用词。的确，现在还不是。我们造出这个词，用来描述你每天碰到的数以吨计的数据。我们撰写了这本书，帮助你在理解小数据时，更得心应手、更灵活、更快捷。

因为哪怕你并没有接受过成为数据专家的相关训练，你依然需要解读数据。如果你是名律师，你需要看出证人是否在陈述证词的时候有所保留。如果你是名护士，你必须理解患者发热的症状是由疾病引起的，还是仅仅与疾病相关。如果你是一

名首席执行官，你会想要预测下个季度收益为多少。如果你是一名家长，你会想要知道自己孩子的能力是否超过（或低于）平均水平。

这本书讲的是如何利用生活中的数据做出更好的决策。

纵观全书，我们会运用真实或假设的例子，重点指出经常被曲解的数据概念。在每一章的结尾，我们会列出 5 种让你运用所学知识的方式。你可以将本书从头读到尾，也可以直接挑你感兴趣的章节读。

本书开篇，记述了脸谱网上两个老友一段简单的留言对话。这段留言的时间为 2013 年秋天，留言双方为约翰（经济学家、统计学家）和迈克（作家）。

“嗨，迈克——我想写本书，脑子里多少有点概念……要不要找个时间聊聊？”

约翰作为职业经济学家，有着同世界各地公司商量如何解读其数据的经验。基于这些经验，他和迈克分享了想要让数据概念变得通俗易懂的想法后，书的脉络渐渐成形。

“明年一整年我们可能会为这件事乐此不疲。”约翰写道。

事实果真如此。（准确说来，是整整两年。）

这条脸谱网留言下的回复一来二去积攒到了 3 288 条。我们期间也发现我们生活中接触的数据比想象中更多。不论何处，只要稍加观察，就能发现越来越多数据被误读、曲解，或者干脆数据本身就是错误的例子。

于是就有了这本书，原本只是在脑中挥之不去的想法，变成了实实在在的文字。

希望你们能喜欢。

在切入正题之前，先呈上一个免责声明：约翰作为职业经济学家和统计学家，他的日常工作包括作为专家证人对数据进行缜密的分析，为《财富》杂志评选出的世界前 100 强公司、贸易集团、政府机构打涉及数百万美元的官司。³ 本书只对关键的数据概念做一个基本的概括性的论述，而非内容全面的教科书。因此如果你反对本书所推崇的让读者在阅读中能“会心一笑”这一特点，那你不用读下去了。有一个好消息，即本书兼具教育性和娱乐性，因此书中肯定有好玩的地方——如解释为什么你汽车的油箱不会完全空掉，为什么名人不会三人同时死亡，以及为什么喝过期牛奶没事。有了上面的免责声明在先，我们现在开始谈谈数据吧。

前言 / IX

序 / XIII

第一章

无处不在的数据：从大数据到小数据

“小数据” / 008

小情境 / 010

成熟的数据接收者 / 011

第二章

对“挑战者号”评估结果的异议：

抽样如何影响结果

1986年1月28日 / 018

理解样本选择 / 019

我们为何需要抽样？ / 023

如果是这样，结果会如何？ / 024

“怪诞”的科学 / 026

抽样不一定越大就越好 / 028

系上安全带 / 029

我们是第1名，也是第58名！ / 031

不接受自拍 / 031

选举总统与人口普查 / 032

取其精华，去其糟粕 / 034

填空 / 035

缺了什么？ / 036

做一个成熟的数据抽样接收者 / 037

第三章

红色州为什么变蓝了：平均数及总数——近观概括性统计

当心数据缝隙 / 047

平均数、中位数和众数哪个更可信？ / 049

迈阿密人出生时平均是西班牙裔，死时是犹太人？ / 052

为什么副市长会比市长赚得多？ / 053

如何评估学生的成绩？ / 056

平均值的平均值 / 057

警惕数据中的异类 / 058

去掉最高分和最低分有道理吗？ / 060

总统办公室的离群值？ / 061

掩盖信息的代价为 10 亿美元 / 062

你是否比一般人更为优秀 / 063

如何成熟对待数据总和、平均值、离群值 / 065

第四章

使用苹果手机的人更聪明？正确理解关联性和因果性

智能手机 = 聪明人？ / 071

星巴克旁边的房子升值更快？ / 073

还有什么因素可以解释这件事？ / 074

我们有没有让你们觉得无聊？ / 077

为什么重要 / 079

穿耐克鞋就能像乔丹一样灌篮？ / 079

婴儿、洗澡水和波尔多红酒 / 081

你在搜索引擎页面排第几？ / 082

烤奶酪的性福生活（我们差点儿将其用作书名） / 083

加利福尼亚的阳光和美属萨摩亚岛上的律师 / 084

不要被媒体人骗了 / 086
 了解大脑的工作模式 / 089
 抛弃先入为主的观念 / 090
 最后但依然重要的一点 / 091
 如何成熟地应对关联性和因果性 / 092

第五章

眼见真的为实吗？我们信仰统计学

民意调查 / 101
 二手烟是否会致癌？ / 102
 重要的事 / 103
 深呼吸 / 105
 抽样大小事关重大 / 106
 你有多大把握？ / 108
 泄露秘密 / 110
 如何面对截然相反的结论 / 112
 效果显著 / 113
 这个研究对我的生活重要吗？ / 114
 等等，还有呢 / 115
 了解自己所看到的是否重要，从而成为成熟的数据接收者 / 116

第六章

非洲为什么会变小？歪曲与曲解

不要臆断 / 124
 柱状图和饼状图应该怎么画？ / 126
 微妙的圆 / 133
 图表让人看起来更值得信任 / 135
 对所有数据一视同仁导致曲解数据 / 136
 体会有和仅有的区别 / 138

油表显示油箱为空为什么还能开个几里路？ / 139
造假与失误 / 139
错误数据抹掉股市 1 360 亿美元 / 141
不要相信维基百科 / 142
过了保质期的食品能不能吃？ / 142
确凿无误的消息也可能被误读 / 144
1/4 磅牛肉汉堡与 1/3 磅牛肉汉堡哪个大？ / 145
如何明智地接收被歪曲（或可能被歪曲）的数据 / 146

第七章

筛选数据需要“摘樱桃法则”：疯狂筛选

筛选数据的“摘樱桃法则” / 156
用统计学的观点如何看祸不单行？ / 158
如何解读体育赛事中的统计数据？ / 160
房价上涨了还是下跌了？ / 161
政客们如何筛选数据？ / 164
相信我们——我们在打广告 / 166
你是“摘樱桃的人”——没错，就是你 / 168
选出最好的，留下其他的 / 168
市场营销人员如何筛选数据？ / 170
如何识别经过筛选的数据？ / 171

第八章

为什么福岛事故是可以预防的：预测未来的技术

明天太阳会照常升起吗？ / 179
预知和预测的区分 / 182
出国旅游之前，请告知金融机构 / 183
正确地看待预测 / 186
抛硬币与“赌徒谬论” / 187

我们身边的预言家 / 190
你知道什么? / 193
民调为什么会出错? / 195
偶然与概率 / 196
心理因素影响预测 / 198
如何成为一个聪明的预测者 / 199

第九章

拨开数据的迷雾：总结

不要轻信头条新闻中的数据 / 204
如何看待工作满意度调查? / 207
出生月份与健康有关系吗? / 209
如何解读关于问题学生的数据? / 212
如何使用房价评估网站的数据? / 214
如何成为一个成熟的数据接收者 / 218
尾声 / 219

词汇表 / 221

注释 / 229

致谢 / 263



