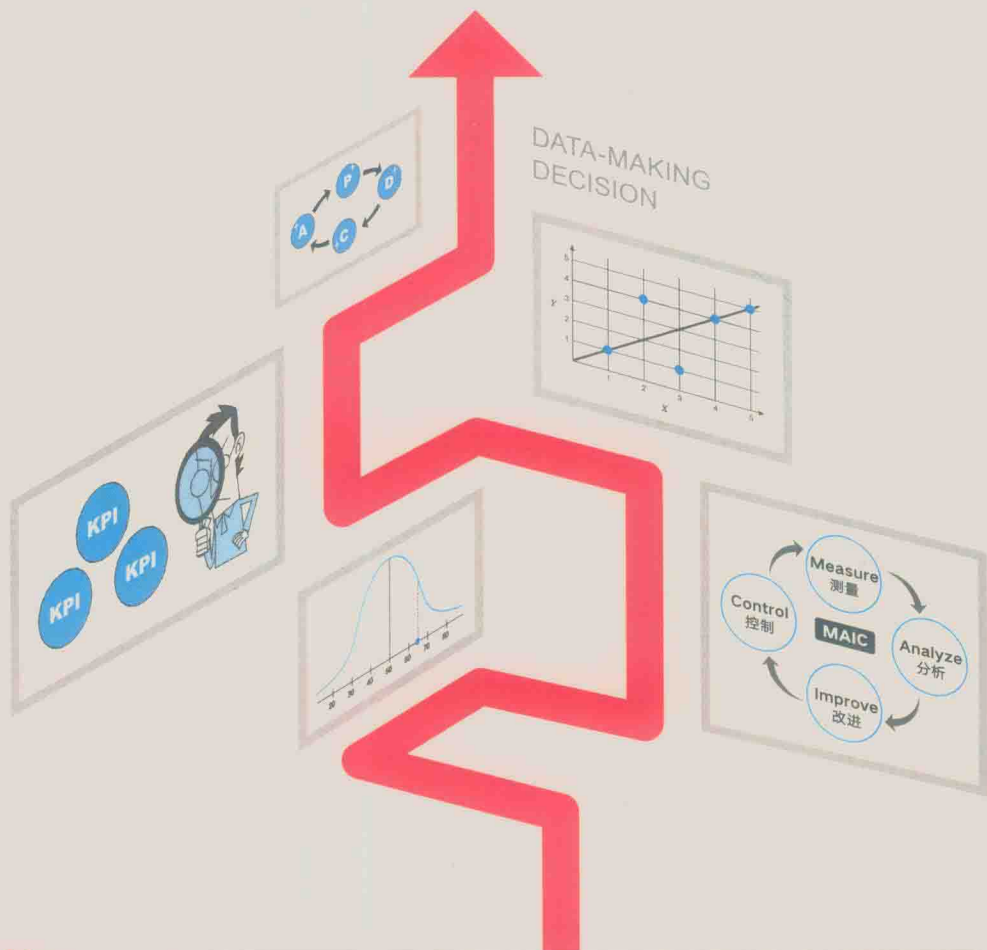


数据化决策

数据分析与高效经营

[日] BSR大数据科学研究会 / 编著

岳冲 / 译



图解版

如何观察市场？如何分析现象？如何预测趋势？如何创造利润？

用数据科学思维 分析商业关键指标

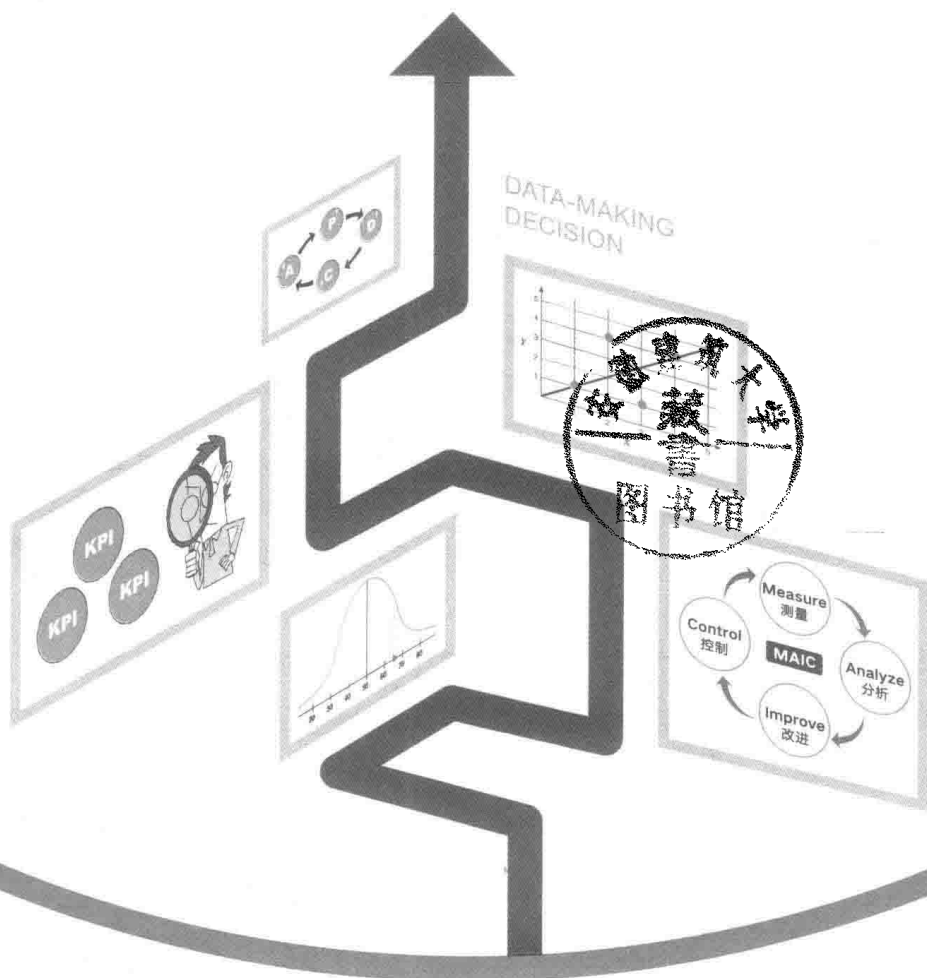
洞悉数据驱动增长的秘密 提升业务导向的决策经营能力

数据化决策

数据分析与高效经营

[日] BSR大数据科学研究会 / 编著

岳冲 / 译



人民邮电出版社

北京

图书在版编目 (C I P) 数据

数据化决策：数据分析与高效经营 / 日本BSR大数据科学研究会编著；岳冲译. — 北京：人民邮电出版社，2018.7

ISBN 978-7-115-47997-6

I. ①数… II. ①日… ②岳… III. ①数据管理
IV. ①TP274

中国版本图书馆CIP数据核字(2018)第044095号

版权声明

“0” KARA HAJIMERU NYUMON DATA SCIENTIST

edited and written by BSR Big Data Science Kenkyukai

Copyright © 2014 Akira Nakano

All rights reserved.

Original Japanese edition published in 2014 by Shuwa System Co., Ltd.

This Simplified Chinese edition published by arrangement with

Shuwa System Co., Ltd., Tokyo in care of Tuttle-Mori Agency, Inc., Tokyo

through Beijing Kareka Consultation Center, Beijing

内 容 提 要

如何用数据分析高效经营已成为职场人士瞩目的焦点。这种现象的背后是社会对从堆积如山的数据中提炼出有益信息的强大需求，可以说正是这种需求使得数据化决策的重要性越来越凸显。

本书围绕数据化决策的方式方法，介绍职场人士所必需的管理知识、统计学基础知识和商务应用能力，并通过 Excel 分析具体案例，深刻阐述从数据分析到高效经营的商务思考方式。

本书不单讲深奥的经营决策思维，也不独论专业的数据分析实操，而是在两者综合的基础上，思维化繁为简，实操具体到位，图文搭配教学，力求帮助职场人士将大数据与商业相结合，找到企业经营之道。

◆ 编 著 [日] BSR 大数据科学研究会

译 岳 冲

责任编辑 恭竟平

责任印制 周昇亮

◆ 人民邮电出版社出版发行 北京市丰台区成寿寺路 11 号

邮编 100164 电子邮件 315@ptpress.com.cn

网址 <http://www.ptpress.com.cn>

北京天宇星印刷厂印刷

◆ 开本：700×1000 1/16

印张：14.5

2018 年 7 月第 1 版

字数：180 千字

2018 年 7 月北京第 1 次印刷

著作权合同登记号 图字：01-2016-8298 号

定价：59.80 元

读者服务热线：(010)81055296 印装质量热线：(010)81055316

反盗版热线：(010)81055315

广告经营许可证：京东工商广登字 20170147 号

企业需要从庞大的数据库中找到有用的信息。

对于企业来说，什么样的信息才有用？

答案是有助于利润提升和企业成长的信息。

这要求每位员工不能只停留在数据分析的阶段，还必须对商业有自己的见解，从而通过数据为企业利润和成长做出贡献。

前言

我好像是一个在海边玩耍的孩子，不时为捡到比通常更光滑的石子或更美丽的贝壳而欢欣鼓舞，而展现在我面前的是完全未探明的真理之海。¹

——艾萨克·牛顿

大数据刚开始轰动世界时，牛顿的这句话浮现在笔者脑海。试将“真理”一词换作“大数据”，是不是也恰如其分地反映了我们当下的状况？

当然，越来越多的人开始认识到浩瀚的大数据的重要性，而试图通过分析、运用大数据寻找真理的活动也变得活跃起来。

与此同时，专业进行数据分析的数据科学家也成为众人瞩目的焦点。这种现象的背后是社会对从堆积如山的数据中提炼出有益信息的强大需求，可以说正是这种需求使得数据分析与商业决策的重要性越来越凸显。

本书聚焦的正是目前备受关注的数据科学家。

当然，在“数据科学家”尚未得到明确定义的当下，记述数据科学家应该具备的知识和技能在某种意义上是一种冒险。但笔者仍然明知山有虎，偏向虎山行，用7章的篇幅描绘了通往数据科学家之路。

¹ 《The Meditation 1978—秋季刊（第五期）》（1978年，平河出版）。

第 1 章是总论，讲述什么是数据科学家、成为数据科学家需要什么样的知识。第 2 章对数据科学家所必需的管理知识进行了阐述。尤其是希望成为数据科学家的 SE（系统工程师），这类管理知识往往是一个短板，但如果能熟练掌握本书中的 KPI 导向思维方式，一切即在掌握之中。此外，本书还借鉴德鲁克的管理理论，提出了一些建议，请将其也作为参考。

在第 3 章到第 6 章中，笔者进一步阐述了数据科学家必须具备的统计学基础知识和商务应用能力。为了让读者深刻理解以数据科学（统计分析）为背景的思考方式，每个要点都列举了使用 Excel 进行数据分析的具体案例。虽然当前统计分析的实际标准正不断向 R 软件靠拢，但读者若无法用 Excel 完成本书所列举的案例，成为数据科学家的梦想也只能是镜花水月。

即使如此，鉴于 R 软件如此受人欢迎，我们仍然应该掌握一些基本知识。最后的第 7 章涉及了一些关于它的内容，为希望进一步提升自己的读者打开了一扇窗。

当然，本书不仅适合立志成为数据科学家的人士，像正文所说，当前普通管理人员也不能缺少数据科学家式的视角，因而他们也适合阅读本书。本书在编排内容时也充分考虑此类人群的需求。

希望更多的读者能够从本书中获取有用的信息。

2014 年 4 月 BSR 大数据科学研究会

目录

第 1 章	大数据时代的数据科学家	1
1.1	点球成金	2
1.2	棒球和数据科学家	4
1.3	发现了“宝藏”的运动家队	6
1.4	在真理的海洋边玩耍的孩子	8
1.5	大数据和数据科学家	10
1.6	大数据时代硬件与网络高度发展	12
1.7	大数据时代另一个特征	14
1.8	最有魅力的数据科学家	16
1.9	数据科学家的素养与技能	18
Management Science ①	什么是管理	20
1.10	数据科学家的用武之地	22
Column 1	直觉判断法	24

第2章	让数据分析服务于经营	25
2.1	三个石匠	26
2.2	数据科学家的眼光	28
2.3	数据科学家的管理素质	30
2.4	战略地图——平衡计分卡	32
2.5	战略地图解读	34
2.6	尺度与评价标准——KPI	36
2.7	KPI 解读	38
	Management Science ② 应当了解的常用 KPI	40
2.8	KPI 导向数据分析	44
2.9	KPI 导向数据分析的基本方式	46
2.10	PDCA 思想及缺陷	48
2.11	约束理论	50
2.12	五大核心步骤	52
2.13	以五大核心步骤实践 PDCA	54
	Column 2 远足比喻	56

第3章 用统计学揭示数据内涵	57
3.1 平均数的内涵	58
3.2 几何平均数及调和平均数	60
Do It Excel ① 用函数计算各类平均数	62
3.3 加权平均数	64
3.4 中位数和众数	66
3.5 频数分布表	68
3.6 直方图	70
Do It Excel ② 绘制直方图	72
3.7 数据分布的离散程度	74
3.8 极差	76
3.9 方差	78
3.10 标准差与变异系数	80
Data Science ① 统计学的关键! 理解方差、标准差、变异系数	82
Do It Excel ③ 计算方差、标准差、变异系数	84
3.11 标准正态分布	86
3.12 Z分数和 T分数	88
3.13 将标准差应用到库存管理中	90
Do It Excel ④ 手头应该持有多少现金?	92
Column 3 6σ (六西格玛)	94

第4章	用相关分析挖掘潜藏商机	95
4.1	相关关系在商业中的重要性	96
	Management Science ③ 理解相关关系和因果关系的不同	98
4.2	验证相关关系的基本步骤	100
4.3	相关关系可视化——散点图	102
4.4	相关关系强弱性——相关系数	104
	Do It Excel ⑤ 散点图和相关系数	106
4.5	问卷调查的相关关系	108
4.6	问卷调查结果的相关矩阵	110
	Do It Excel ⑥ 用分析工具计算相关系数	112
4.7	满意度分析图	114
4.8	策略方案的制定	116
	Do It Excel ⑦ 用散点图绘制满意度分析图	118
4.9	顾客与商品的细分化	120
	Do It Excel ⑧ 绘制帕累托图	122
4.10	交叉分析的基本方法	124
	Do It Excel ⑨ 用数据透视表进行交叉分析	126
	Column 4 找不到“数据分析”按钮!	128

第5章 用回归分析预测未来	129
5.1 用回归分析预测未来	130
5.2 回归直线与回归方程	132
5.3 回归直线准确度—— R^2 值	134
Data Science ② 计算 R^2 值的思考方式	136
Do It Excel ⑩ 计算回归系数和截距	138
5.4 理解回归分析及其结果的含义①	140
5.5 理解回归分析及其结果的含义②	142
Do It Excel ⑪ 用回归分析表导出回归方程	144
5.6 多元回归分析	146
5.7 用 Excel 进行多元回归分析	148
5.8 多元回归方程	150
5.9 报纸广告和店内促销，谁的贡献度更高？	152
5.10 定性转定量，分析更简单①	154
5.11 定性转定量，分析更简单②	156
Do It Excel ⑫ 将有、无转为数值	158
5.12 回归分析表	160
5.13 用回归分析表进行有效促销	162
Column 5 看似简单实际很难的概率问题	164

第6章 用检验做出战略决策	165
6.1 假设是否成立?	166
6.2 样本均值与总体均值	168
Data Science ③ 中心极限定理	170
6.3 通过掷骰子理解检验	172
Do It Excel 13 制作二项分布图	174
6.4 检验的基本步骤	176
6.5 该店午餐价格比当地午餐均价便宜?	178
Data Science ④ 双边检验和单边检验	180
Do It Excel 14 调查平均数之间是否存在差异①	182
6.6 样本容量小? —— t 检验	184
Do It Excel 15 调查平均数之间是否存在差异②	186
6.7 多个比较对象? —— F 检验	188
6.8 是否有显著差异? —— 方差分析①	190
6.9 是否有显著差异? —— 方差分析②	192
Column 6 均值回归	194

第7章 用 R 语言结构化数据	195
7.1 数据分析的实际标准——R	196
7.2 R 的基本操作	198
Do It R ① 读取 R 的数据集	200
7.3 箱线图	202
Data Science ⑤ 理解箱线图的内涵	204
Do It R ② 导入程序包	206
7.4 用 R 也能做回归分析	208
7.5 用 R 还能做聚类分析	210
Do It R ③ 用 R 进行聚类分析	212
Do It R ④ 将城市间的距离可视化	214
7.6 小心黑天鹅事件	216
参考文献	218

数据化决策：数据分析与高效经营

第 1 章

大数据时代的数据科学家

大数据时代，我们每个人都淹没在大量的信息当中。

要从爆炸式增长的数据中找出有效信息，必须借助“专家”的力量。

这些“专家”就是数据科学家。

本章试图解读数据科学家越来越受重视的背景以及如何成为一名数据科学家。

1.1

点球成金

棒球史上的真实事件

你听说过《点球成金》吗？

估计很多人会回答说看过电影或读过原著¹。《点球成金》的故事取材于美国职业棒球联盟奥克兰运动家队的一段真实历史。

比利·比恩（Billy Beane）曾是一名备受期待的棒球选手，但他始终未能在成绩上有所突破，不久后他选择去曾经的豪门——运动家队做总经理。在球队经营上，比恩和其他球队领导最大的不同点就是他特别重视数据。

单是这样也并不稀奇，日本的野村教练也非常擅长分析数据。比恩最独特的地方在于他选择数据的视角。

重视不同的标准

在比恩看来，以往的棒球都过于夸大了选手安打率（击球率）、奔跑速度、守备技能及身体素质的重要性。相反，控制好球区的能力才是决定一名棒球选手能否走向成功的最大因素。

因此比恩在分析数据时并没有将目光投向选手的安打率，而是选择了上垒率。

上垒率和球队得分的相关系数达到了 0.92，超过了安打率和得分间的相关系数。

基于这种指导思想，比恩开始打造一支重视上垒率的球队。

比恩的尝试是对安打率至上的传统球队建设方式的一种挑战。

¹ 迈克尔·刘易斯著，中山宥译《点球成金》（2004年，兰登书屋讲谈社）。

● 奥克兰运动家队和比利·比恩总经理

棒球中最重要的
是选手的安打率
和奔跑速度、
高超的守备技能
及身体素质。



以往的球队总经理

不，
棒球中最重要的
是选手的上垒率！



奥克兰运动家队的
总经理比利·比恩

奥克兰运动家队



加利福尼亚州奥克兰的美联球队，属于美联西区。2013 年获得美联西区冠军。

● 比利·比恩 (1962~)

作为选手进入美联，不久后就任运动家队总经理。目前仍担任总经理，带领队伍不断前进。



* 图片来自：GabboT

棒球和数据科学家

上垒率 $\times 3$ + 长打率

1997年比恩就任总经理后，立刻对球队经营进行了改革。改革大棒首先挥到了球探的头上。

过去，球探一直是凭借自己的感觉和经验为球队发掘有发展潜力的选手。比恩上任后对这种方式进行了改革。他将目光集中在选手的上垒率方面，以此为标准来筛选新人。这遭到了球探们的强烈反对，他们甚至陆续辞职。

此外，为了更好地推进重视数据的球队建设方式，比恩还在1999年邀请保罗·德波戴斯塔进入球队管理层。

德波戴斯塔毕业于哈佛，曾将棒球比赛获胜的必要因素作为研究课题。他把20世纪所有球队的数据代入算式中，试图找出胜利的方程式¹。他就是我们现在所说的数据科学家。

德波戴斯塔敏锐地察觉到棒球中最重要的数据只有两项——上垒率和长打率。他在加入运动家队之后又进行了更深一步的思考——上垒率和长打率的重要程度是否相同？

针对这个疑问，德波戴斯塔在进行数据分析后得出结论：上垒率的重要程度若是3，长打率则为1。为此，运动家队设定了一个至关重要的指标。

$$\text{上垒率} \times 3 + \text{长打率}$$

¹ 迈克尔·刘易斯，《点球成金》P167。