

深入浅出 统计学

轻松剔除
图形错误



让统计概念
植根于大脑



避免难堪的
抽样失误

利用标准差提高
赛季平均成绩



赌场趋吉避凶



看统计学如何
瞒天过海

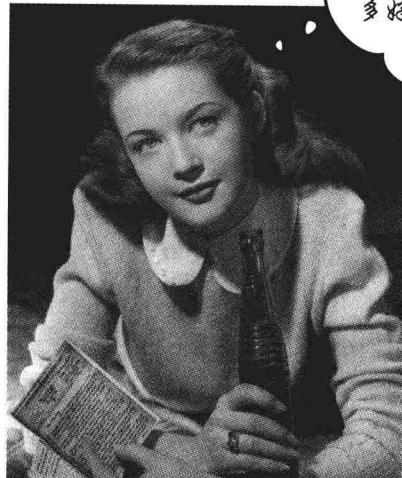
Dawn Griffiths 著
李芳 译

O'REILLY®

深入浅出统计学

Head First Statistics

要是有那么一本关于统计学的书，不再像看牙医那么恐怖，该有多好？可这不过是个梦罢了……



Dawn Griffiths 著
李芳 译

电子工业出版社

Publishing House of Electronics Industry

北京 • BEIJING

内容简介

《深入浅出统计学》具有深入浅出系列的一贯特色，提供最符合直觉的理解方式，让统计理论的学习既有趣又自然。从应对考试到解决实际问题，无论你是学生还是数据分析师，都能从中受益。本书涵盖的知识点包括：信息可视化、概率计算、几何分布、二项分布及泊松分布、正态分布、统计抽样、置信区间的构建、假设检验、卡方分布、相关与回归等等，完整涵盖AP考试范围。本书运用充满互动性的实际世界情节，教给你有关这门学科的所有基础，为这个枯燥的领域带来鲜活的乐趣，不仅让你充分掌握统计学的要义，更会告诉你如何将统计理论应用到日常生活中。

978-0-596-52758-7 Head First Statistics © 2009 by O'Reilly Media, Inc. Simplified Chinese edition, jointly published by O'Reilly Media ,Inc. and Publishing House of Electronics Industry, 2011.Authorized translation of the English edition, 2009 O'Reilly Media, Inc., the owner of all rights to publish and sell the same.All rights reserved including the rights of reproduction in whole or in part in any form.

本书中文简体版专有版权由 O'Reilly Media, Inc. 授予电子工业出版社，未经许可，不得以任何方式复制或抄袭本书的任何部分。

版权贸易合同登记号 图字：01-2011-7144

图书在版编目（CIP）数据

深入浅出统计学 / (美) 格里菲思 (Griffiths,D.) 著；李芳译. — 北京：电子工业出版社，
2012.1 书名原文：Head First Statistics

ISBN 978-7-121-15308-2

I . ①深… II . ①格… ②李… III . ①统计学—通俗读物 IV . ① C8-49

中国版本图书馆 CIP 数据核字 (2011) 第 243859 号

责任编辑：李影 策划：卢鹤翔

印 刷：北京天宇星印刷厂

装 订：三河市鹏成印业有限公司

出版发行：电子工业出版社

北京市海淀区万寿路 173 信箱 邮 编：100036

开 本：860×1092 1/16 印 张：45 字 数：717 千字

印 次：2012 年 1 月第 1 次印刷

印 数：4 000 册 定 价：89.00 元

凡所购买电子工业出版社图书有缺损问题，请向购买书店调换。若书店售缺，请与本社发行部联系，
联系及邮购电话：(010) 88254888。

质量投诉请发邮件至 zlts@phei.com.cn，盗版侵权举报请发邮件至 dbqq@phei.com.cn。服务热线：
(010) 88258888。

“《深入浅出统计学》是目前市面上最具娱乐性、最能够抓住读者注意力的统计学研读指南。透过生动活泼的手法与素材，为这个困难的主题提供最容易被接受的学习方式，贯穿全书的精辟解说让各种程度的学生都能够充分地理解统计学的妙义。”

—— 阿瑞娜·安德森（Ariana Anderson），加利福尼亚大学洛杉矶分校统计系教师助理及博士生

“《深入浅出统计学》润物细无声。当一口气看完讲解和练习后，你就会发现自己在社交谈话中可以开口闭口正态分布、泊松分布，我保证并没有人建议你这么做！”

—— 加里·沃尔夫（Gary Wolf），《连线》杂志（Wired Magazine）特约编辑

“道恩·格里菲思把一些十分复杂的概念拆分为一块块小材料，它们不那么令人望而生畏，凡夫俗子都会觉得十分容易掌握。大量图形、图片让材料具体生动，458页那位吵着要买口香糖球的迷人女模特已然让我心生情愫。”

—— 布鲁斯·弗雷（Bruce Frey），《统计学技巧》（Statistics Hacks）作者

“《深入浅出统计学》运用简单的生活实例，提供最符合直觉的理解方式，让统计理论的学习既有趣又自然。”

—— 迈克尔·普瑞诺（Michael Prerau），波士顿大学计算神经科学和统计学讲师

“你以为‘深入浅出’图书只适合计算机迷吗？不妨试试用本书提供的学习方式学习统计学，你就会改变想法。这方法的确有用。”

—— 安迪·帕克（Andy Parker）

“这本书非常适合学生学习统计学——寓教于乐、讲解全面、易于理解。完美无缺的方法！”

—— 丹妮尔·莱维特（Danielle Levitt）

“打倒其他枯燥无味的统计书！连我的猫都喜欢这一本。”

—— 凯里·科利特（Cary Collett）

深入浅出系列图书美誉

“凯西和伯特合著的《深入浅出Java》让白纸黑字摇身一变，成为读者领略过的最接近图形用户界面的作品。作者以幽默、新潮的风格，让学习Java成为不断追问‘他们后来怎么样了?’的愉快体验。”

——沃伦·柯费尔（Warren Keuffel），《软件开发》杂志（Software Development）

“《深入浅出Java》引人入胜的风格会把本来一无所知的你变成斗志昂扬的Java战士，不仅如此，书中还收入了大量实用事例，这样的实用事例在其他文章中只会留给恐怖的‘读者练习’。此书睿智、幽默、新潮而实用——能在讲授对象序列化和网络加载协议知识的同时有这样的主张并坚持做到的书籍并不多见。

——丹·罗素（Dan Russell）博士，IBM Almaden研究中心用户科学和用户体验研究室主任（兼斯坦福大学人工智能教师）

“此书明快，风趣，玩世不恭，引人入胜。细心读——你可能确实能学到东西！”

——肯·阿诺德（Ken Arnold），曾任Sun Microsystems高级工程师，与Java创始人詹姆斯·戈斯林（James Gosling）合著《Java编程语言》（The Java Programming Language）

“如醍醐灌顶，脑海中堆积如山的书本知识一下子消化了。”

——沃德·坎宁安（Ward Cunningham），维基概念发明人，Hillside Group创立人

“正合我们这些喜欢研究技术、生活随意的程序员的口味，实用开发策略的称手参考书——让我的大脑尽情运转，无须硬着头皮应付迂腐乏味的专家说教。”

——特拉维斯·卡兰尼克（Travis Kalanick），Scour网站和Red Swoosh网站创始人，获麻省理工学院TR100（《技术回顾》世界百名青年创新学者）称号

“有的书是用来买的，有的书是用来藏的，还有的书是用来摆在案头的。感谢O'Reilly和Head First的员工，他们出了最高等级的书——深入浅出系列，让人爱不释手、百读不厌。《深入浅出SQL》是我最心爱的书，连参考用的PDF打印稿都快翻烂了。”

——比尔·索耶（Bill Sawyer），Oracle公司ATG课程经理

“本书的透彻、幽默和睿智令人钦佩，连编程门外汉也能借助这样的书想出办法解决问题。”

——科利·多克托罗（Cory Doctorow），博客网站BoingBoing撰稿人合作编辑，著有《魔法王国的故事》（Down and Out in the Magic Kingdom）及《人来人往的城市》（Someone Comes to Town, Someone Leaves Town）

“昨天收到书就开始读……一读就停不下来了，真是酷毙了。书很有趣，内容扎实，切中肯綮。印象太好了。”

——艾瑞克·伽玛（Erich Gamma），IBM杰出工程师，《设计模式》（Design Patterns）合著者

“我读过的最有趣、最高明的软件设计图书之一。”

——艾伦·拉伯奇（Aaron LaBerge），ESPN.com技术副主席

“过去要犯着错误摸索前进的漫长学习过程，现在干净利落地浓缩在一本迷人的平装书中。”

——麦克·戴维森（Mike Davidson），Newsvine, Inc.首席执行官

“每一章都凝聚着优雅的设计，每一条原理无不饱含实用价值与闪光智慧。”

——肯·戈德斯坦（Ken Goldstein），迪士尼在线执行副总裁

“我 ♥ 《深入浅出HTML与CSS & XHTML》。它以‘有趣’的模式，将全部知识倾囊相授。”

——丹妮尔·莱维特（Danielle Levitt）

“通常，阅读设计模式方面的书或文章时，我都得头悬梁锥刺股才能保证注意力集中。这本书却是个例外，听起来可能有点怪，这本书让学习设计模式变得盎然有趣。

“当其他设计模式方面的书籍还在教读者呀呀学语时，这本书却已在踏浪高歌‘加油，兄弟！’”

——艾维克·伍勒（Eric Wuehler）

“我实实在在爱这本书。事实上，我当着老婆的面亲了这本书。”

——塞提斯·库玛（Satish Kumar）

O'Reilly其他相关图书

Statistics Hacks™

Statistics in a Nutshell

Mind Hacks™

Mind Performance Hacks™

Your Brain: The Missing Manual

O'Reilly深入浅出系列其他图书

Head First Java™

Head First Object-Oriented Analysis and Design (OOA&D)

Head First HTML with CSS and XHTML

Head First Design Patterns

Head First Servlets and JSP

Head First EJB

Head First PMP

Head First SQL

Head First Programming (2008)

Head First Design Patterns

Head First Servlets and JSP

Head First EJB

Head First PMP

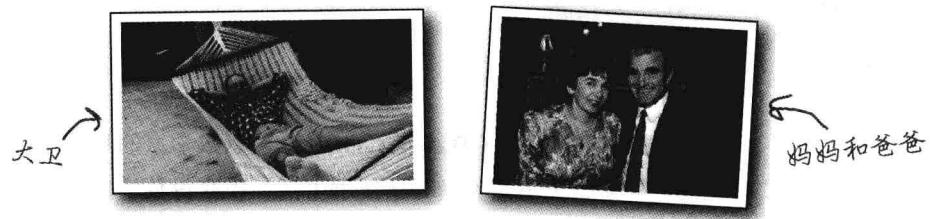
Head First SQL

Head First Programming (2008)

Head First PHP & MySQL (2008)

Head First Algebra (2008)

献给我的爸爸、妈妈、大卫和卡尔。谢谢你们的支持和信任。
不过还要等一段时间才能给你们车！



作者介绍



道恩·格里菲思
(Dawn Griffiths)

道恩·格里菲思（Dawn Griffiths）以数学家的身份在一所顶尖英国大学开始读书生活，曾获得数学专业一级荣誉学位。可是，当她意识到人们在聚会上不愿意与她谈话时，就拒绝了一份研究极为罕见的微分方程的博士奖学金，转而从事软件开发事业。目前她将IT咨询、写作及数学集于一身。

当道恩不写深入浅出系列图书的时候，你会看到她在打太极拳、织梭结花边或在烹饪美食——她还没有修炼出一心三用的绝技。旅行以及和她深爱的丈夫大卫在一起消磨时光也让她十分享受。

道恩有一个猜测——《深入浅出梭结花边》有可能会很火爆，可她怀疑布雷特（Brett）和劳里（Laurie）不会同意。

译者序

《深入浅出统计学》翻译好了。这是我翻译的第二本“深入浅出”系列书籍，第一本是《深入浅出数据分析》，两本书可以说是姊妹篇。

《深入浅出统计学》具有深入浅出系列的一贯特色，即“认为深入浅出系列图书的每一位读者都是一位学习者”，遵守“将知识图形化”、“采用对话式的个性化风格”、“引导读者深入思考”、“牢牢吸引读者的注意力”、“影响读者的情感”等约定。不过，这并不意味着无法捕捉到作者的个人特点。

与《深入浅出数据分析》喧嚣错杂的现场气氛相比，《深入浅出统计学》显示出一分属于“办公室”的宁静，由数据分析师整理好的数据就摆在那里，找到合适的方法进行演算成为工作的重点，这些方法包括：信息图形化、集中趋势的量度、分散性与变异性的量度、概率计算、离散概率分布、排列与组合、几何分布、二项分布及泊松分布、正态分布、统计抽样、总体和样本的估计、置信区间的构建、假设检验、卡方分布、相关与回归。从开篇至收尾，这些方法由独自登场到联袂表演，从“一支红杏出墙来”到“乱花渐欲迷人眼”，徐徐揭开统计学神秘的面纱。作者充分利用几个活泼有趣的情境，淋漓尽致地发挥深入浅出系列“反复阐述”的特色，举重若轻、有条不紊地完成了这本篇幅长达717页、内容多达15章的作品。

身为译者，我当然有责任全力体现这样一部作品的风采。只是怎样一本译作才是“好”译作？这个议论伴随翻译这个营生的生而生，也许也只能伴随翻译这个营生的亡才会亡。与其孜孜不倦地讨论，不如做点切实的事情，向着“好”的目标接近。

在《深入浅出统计学》的翻译过程中，有两件工作可以斗胆请读者放心：

一是术语的翻译。所有术语都有合理出处，优先采用的参考资料是《GBT 3358.1—2009 统计学词汇及符号 第1部分：一般统计术语与用于概率的术语》。为了方便读者研究，现已将此标准中的所有术语整理成《GB术语表》，见此链接：<http://t.cn/aO8vlq>。除了少数几个例外，只要能在此标准中查到的术语，均优先使用。所提到的少数几个例外则是为了成全表述上的一致性和逻辑性，例如，“range”一词，未采用GB中出现过的“极差”，而是采用其他资料中出现的另一种说法“全距”，这是为了让与“range”有关的几个术语——全距、四分位距、百分位距等保持和谐一致，便于理解和记忆。

二是标点符号和字体格式的认真使用：第一，统计术语首次出现时加粗，凸显这个术语，引起读者重视；第二，统计用语过于冗长时，为了避免这个用语与句子的其他组成部分混淆不清，也加引号进行提示，例如，第4章第133页上的“停球结果为7”；第三，提高破折号的使用频率，当前后两个句子有较为密切的因果联系时，即使原文未使用破折号，译文也改句号为破折号，以作提示。

遗憾的是，有极少量术语无法做到既坚持采用术语，又保持中文的流畅。这固然是由于我个人经验不足、能力有限，但中英文的习惯差异、技术类书籍对术语正确性和一致性的特定要求也是其中的原因。例如，“expect”译为“期望”，可是，当读到第5章中“这是你每拉一次杆能够期望得到的收益——是负数！”这样的句子时，我们多少会觉得有些怪异——在中文中，我们习惯于将“期望”这个词和美好的前景联系在一起，谁会期望自己的收益是负数？可是，如果将“期望”顺手改为“预期”以满足流畅性的需要，那么从技术的严谨性来看，并不是十分负责的做法。这样的词还有“成功”——只要某件事情如期发生，都叫做“成功”，无论好歹。每到这种纠结时分，我往往按照个人意愿，任性地坚持“直译”。衷心希望这样的不足不会影响读者对统计学的兴趣和研究。

本书初稿提交后，在卢鹤翔编辑的热心联络下，以下素未谋面的专业人士参与了初稿试读，给出了宝贵的审定意见，现借译者序向以下热心人士致谢：

陈钢、高涛、江思源、梁涛、林宇、刘思喆、卢涛、孙强宇、魏太云、吴昕、徐旭铭、张伸

最后：

感谢电子工业出版社、武汉嘉士特文化传播公司对我的信任。

感谢卢鹤翔、徐涛编辑对我的帮助和指教。

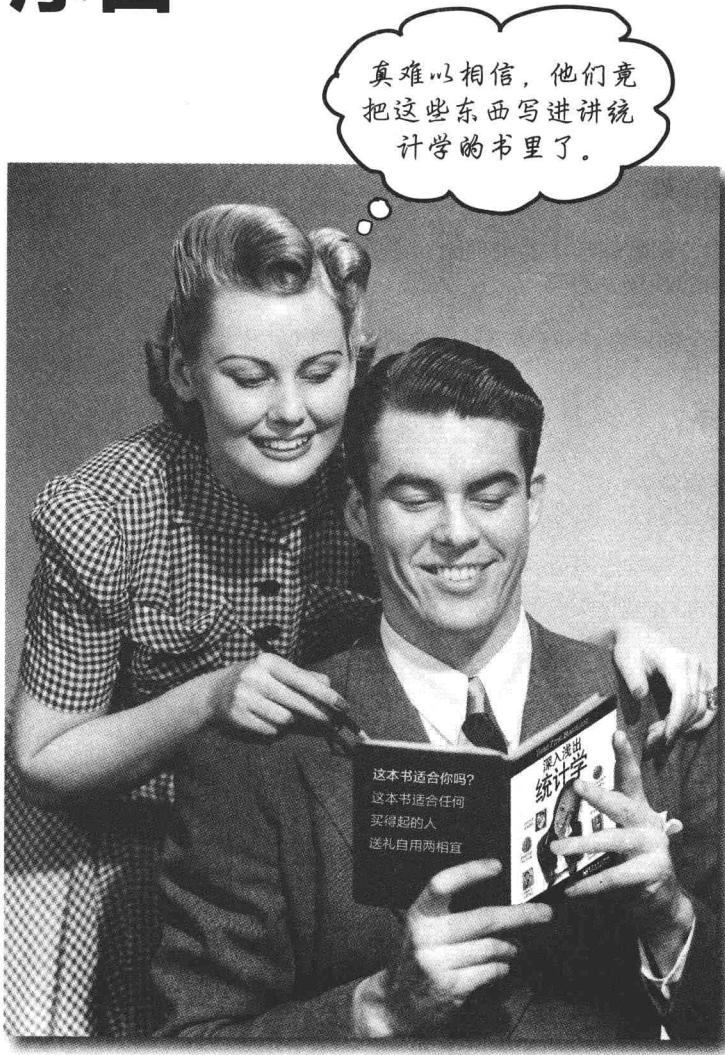
感谢家人对我的理解和支持，我爱你们。

李芳

2011年9月

如何使用本书

序言



本节回答一个热门问题：“作者为什么要
把这些东西写进一本讲统计学的书里？”

谁适合阅读本书？

请先回答几个问题：

- ① 因为一门课程、因为自己的工作，或仅仅因为你觉得自己应该懂得标准差的意义或是如何求轮盘赌赢钱概率，使你需要了解统计学？
- ② 你想学习、理解和记忆如何使用概率和统计百发百中地得出正确结果？
- ③ 与枯燥、无聊的学术演讲相比，你更喜欢参加充满欢声笑语的晚宴，对吗？

如果以上问题全部回答“对！”——这本书适合你。

谁该和本书说拜拜？

请先回答几个问题：

- ① 你从来没有学过基础算术？（不用很精通，但应该会基本的加、减、乘、除。）
- ② 你是一位意气风发的统计师，正在物色参考书？
- ③ 你惧怕尝试新事物，宁可上山打虎也不愿标新立异，对吗？你认为要是用拟人的手法叙述维恩图，统计学书籍就难免有失严肃，对吗？

只要有一个问题回答“对！”
—— 你与本书无缘。



[营销部捎话——只要有信用卡就可以买书哦。]

我们了解你在想什么

“这怎么能是一本严肃的统计学书籍呢？”

“这些图都是用来干嘛的？”

“我真能这样学统计学吗？”

我们了解你的大脑在想什么

你的大脑渴望新事物。大脑总是不停地搜索、探查、等待不同寻常的事物，它天生如此，这正是你活力的来源。

那么，大脑怎么对待你所碰到的常规、普通、一般的事情呢？——它会竭尽全力阻止这些事情，以免干扰自己真正的工作——记录重要事项。大脑不会费力保存这些琐事，这些琐事从来不会成功地闯过“明显不重要事项”关卡。

你的大脑如何知道哪件事重要？假想有一天你出门旅行，迎面扑来一只吊睛白额大虫，你的头脑和身体会有什么反应？

神经元发动、情绪激动、化学物质激增。于是，你的大脑知道——

这事绝对重要！记住！

但，想象你是呆在家里，或是呆在图书馆里，也就是说，是在一个安全、温暖、没有老虎的地方。

你正在复习迎考，要不然就是在努力搞懂一些艰深的技术，你的老板认为花个把星期就能搞定，顶多十天。

唯一的问题是：你的大脑想好好帮你一把，它试图保证不让这种“明显不重要”的内容去破坏珍惜的资源，这些珍稀的资源最好用来保存真正“重大”的事情，像老虎啊，像火灾险情啊，像你绝不该在Facebook的网页上贴上那些“聚会”照片啊。

没有什么方便的办法可以告诉大脑“喂，大脑，我对你感激之至，可是不管这本书多无聊，也不管我对这本书多没兴趣，我真的想把这些东西全都记住。”



我们认为深入浅出 (Head First) 系列 图书的每一位读者都是一位学习者。

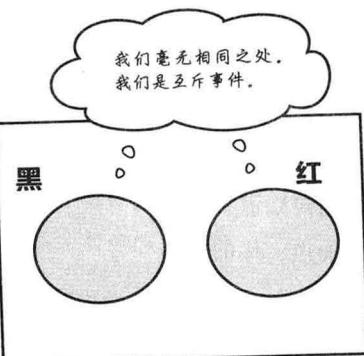
既然要学习，怎样才能学会呢？首先，你得搞懂，然后，切勿遗忘。一字一句硬塞不是办法。根据最新的认知科学、神经生物学及教育心理学研究结果，学习远不仅仅是读书认字。我们知道怎么让你的脑筋动起来。

下面是部分深入浅出 (Head First) 教学原则：

将知识图形化。图形比单调的文字好记得多，可以提高学习效率（记忆学习和转移学习的学习效率最多能提高89%）；图形还能让知识更容易理解，相比将文字放在页脚和下一页，**将文字放在相关图形当中或图形周围**，学习者成功解决相关问题的可能性将成倍增长。



采用对话式的个性化风格。最近的研究表明，要是回避一本正经的语气，代之以对话般的风格，以第一人称平易近人地给学生上课，学生的课后测验成绩最多可提高40%。多讲几个故事，少来一点高谈阔论，语气宜随和。别太郑重其事。想想看，一局欢声笑语的晚宴，或是一场枯燥的演讲，哪一样更让你感兴趣？



s

引导读者深入思考：换句话说，除非读者主动调动自己的神经元，否则脑袋里不会发生什么大变化。只有激发读者的兴趣，引起读者的好奇，刺激读者的灵感，读者才能解决问题，得出结论，获得新知识。为此，讲授者要设计各种难题、练习，提出引人深思的提问，还要多设计一些让左右脑半球和多种感官都动起来的活动。



牢牢吸引读者的注意力。大家都有这样的体验——“我是真想学，但看完第一页就晕了”。大脑注意的是不同寻常的、有趣的、奇怪的、引人注意的、出人意料的事情。学习一种新颖艰深的技术不一定非得枯燥乏味不可，果真有吸引注意力的内容，大脑会学得更快。

影响读者的情感。现在我们知道，人的记忆能力在很大程度上取决于要记忆的内容对情感的影响。我们关心什么，就会记住什么；我们对什么事有感觉，就会记住什么。这里讲的情感并非天灾人祸给人带来的撕心裂肺的伤痛情感，而是惊讶、好奇、感觉有趣、想追根究底之类的情感，以及在猜对一个字谜、在学会别人感觉难以学会的事情或是在意识到自己懂的东西居然比工程部那位开口闭口“我比你有技术”的张三还多时，油然而生的“我是老大”的感觉。



元认知：对思考的思考

如果真想学东西，而且想学得更快更深入，就要关注自己如何集中注意力。要思考自己的思考方式，研究自己的研究方式。

大多数人在成长过程中都不曾学习元认知和学习理论方面的知识。人们期望我们学知识，但极少有人教我们如何学。

但想象得到，捧着本书的你，的确想学习统计学知识，同时可能不想花费太多时间。要想利用在本书中读到的知识，就得记住读过的知识，为此必须理解这些知识。为了淋漓尽致地发挥本书或任何书本或学习经验的作用，请管好你的大脑，请管好大脑对待本书的态度。

诀窍在于让大脑把正在学习的新资料当做“正经大事”——对幸福至关重要的大事，像面对老虎一样重要的大事。若非如此，你就会陷入一场持久战：你竭力要记住新知识，大脑却竭力要把这些新知识踢出去。

既然如此，如何让大脑像对待吃人的老虎一样对待统计学知识呢？

有两种办法，一种缓慢而乏味，一种迅速而有效。慢办法是简单记忆。你显然明白，只要不停地把同样的东西往大脑里灌，即使是最乏味的知识，也能学会、记牢。只要重复灌的次数足够多，大脑就会想：“这些东西给他的感觉并不重要，但他不停地看这些相同的东西，一遍，一遍，再一遍。因此我猜这些东西肯定很重要。”

快办法是做一切增进大脑活动的事，尤其是不同类型的大脑活动。上一页讲了很多这样的活动，事实证明，这些活动全都能促使大脑以有利于己的方式工作。例如，研究表明，将文字放在文字所描述的图片当中（相反的做法是将文字放在页面中的其他位置，如注释位置或正文位置），会促使大脑努力想搞清楚文字和图片之间的关系，进而发动更多神经元。更多神经元发动 = 更有机会让大脑明白某件事值得注意，可能还值得记住。

对话式的写作风格对此很有帮助。人们在与人对话时注意力会更集中，原因是别人期待对方有所表现。令人惊讶的是，大脑不一定会在意“对话”是在人和书之间进行！反之，要是写作风格了无新意，乏味枯燥，大脑的感觉就和在挤满消极听众的屋子里听演讲没什么两样：没必要保持清醒。

不过，图形和对话式风格只是起步……



我们的做法：

我们使用丰富的图片，这是因为，大脑追逐图像，而非文字。在大脑的活动中，一张图片胜过千言万语。当同时使用图片和文字进行说明时，我们将文字填写在图片当中，当文字出现在它所描述的事物当中时，大脑的工作更有效率；相反，若将说明性文字放在注释或其他正文当中，则无此效果。

我们使用**反复论述法**，即以不同的方式、通过不同的媒介对同一主题进行反复描述，给读者营造丰富的感受，目的是让这些主题有更多机会印在大脑的多个区域。

我们以**出人意料**的方式叙述概念和使用图片，因为，大脑追逐新鲜事物；我们在图片和创意中或多或少加入了一些**情感性**的内容，因为，大脑关注情感的生物化学反应。让人有所感触的东西更可能让人记住，即使这点感触不过是一丝幽默、一丝惊讶或一丝兴趣，也是如此。

我们使用**个性化的对话式写作风格**，因为，当大脑认为你是在进行对话而不是在消极地听报告时，就会调整到注意力更集中的状态。即使在读书时，大脑也是这个习惯。

我们安排了**80多个活动**，因为，相比读书，在做事时，大脑经过调整，能学会和记住更多东西。我们安排的练习有难度，但不会让人束手无策，这正是大多数人愿意做的练习。

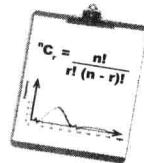
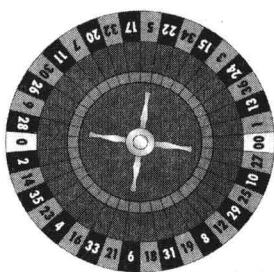
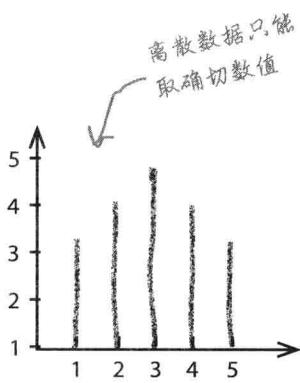
我们使用**多种教学风格**，因为，有的人可能喜欢一步一步按顺序来，有的人可能喜欢先看懂大图，还有一些人可能只想看看例子。我们将以多种方式反复讲述相同的主题，不管读者的个人爱好如何，都将因此受益匪浅。

我们安排了让**左右脑半球**分别负责的内容，因为，大脑开动部位越多，就学得越多，记得越多，注意力更持久。由于一侧大脑工作往往意味着另一侧大脑得到休息，左右半脑的分工合作使得长时间学习的学习效率得到提高。

我们还安排了一些**场景**和练习，在场景中展现**不同的观点**，因为，当大脑被迫进行评估和判断时，会调整到深入学习状态。

我们在练习中安排了一些**难点**，即提出一些无法简单回答的问题。因为，你的大脑在不得不处理某件事情时，会调整到学习和记忆状态。开动脑筋吧，“光看别人做运动无法让自己体态健美”。别担心，我们尽力保证，你努力学习的都是该学的，你不会为了对付一个费解的例子或为了分析一段用词过于晦涩或行文过于简化的段落而多用一个脑细胞。

我们以**人物**为例，把人物安排在场景、实例、图片等等内容中。至于原因嘛，因为你是人群中的一个啊，你的大脑对人比对事更关注。



重要统计量

