世界的每一个角落，生活的每一个细节，统计学无所不在。

趣味图解统计学知识，谁说菜鸟学不会数据分析？

（日）**小岛宽之**◎著　孔霈◎译

你一定爱读的极简统计学

再精简下去，就不是统计学了！

清华大学副教授/**肖勇波**推荐！

日本**热印18次**，零基础入门读物！

台海出版社

你一定爱读的
极简统计学

（日）小岛宽之◎著　孔霈◎译

再精简下去，就不是统计学了！

台海出版社

图书在版编目（CIP）数据

你一定爱读的极简统计学：再精简下去，就不是统计学了/（日）小岛宽之著；孔霈译. -- 北京：台海出版社，2014.4

ISBN 978-7-5168-0451-3

Ⅰ. ①你… Ⅱ. ①小… ②孔… Ⅲ. ①统计学－基本知识 Ⅳ. ①C8

中国版本图书馆CIP数据核字（2014）第210577号

著作权合同登记号　图字：01-2014-5354

你一定爱读的极简统计学：再精简下去，就不是统计学了

著　　者：（日）小岛宽之　　译　　者：孔　霈

责任编辑：王　艳　　装帧设计：久品轩

版式设计：刘丽娟　　责任印制：蔡　旭

出版发行：台海出版社

地　址：北京市朝阳区劲松南路1号，　邮政编码：　100021

电　话：　010－64041652（发行，邮购）

传　真：　010－84045799（总编室）

网　址：　www.taimeng.org.cn/thcbs/default.htm

E-mail：　thcbs@126.com

经　销：全国各地新华书店

印　刷：北京彩虹伟业印刷有限公司

本书如有破损、缺页、装订错误，请与本社联系调换

开　本：170×230　1/16

字　数：145千字　　印　张：13.25

版　次：2015年1月第1版　　印　次：2015年1月第1次印刷

书　号：ISBN 978-7-5168-0451-3

定　价：38.00元

推荐序

让统计学成为一种大众数据分析工具

清华大学经济管理学院 肖勇波 副教授

在信息技术高度发达的今天，电子商务、社交网络、云计算等新兴应用已经渗透到人们生活的每一个角落，我们正在进入一个大数据的时代。如何有效发挥出大数据的价值？如何有效应对大数据带来的挑战？这恐怕是当今世界各行各业最为关心的问题之一。诚然，大数据的价值并不在于数据本身，而在于隐藏在数据背后的有用的知识。如何发现这些看不见的知识？这是一个仁者见仁、智者见智的事情，也是目前企业界和学术界正在全力探索的问题。

大数据的商务分析貌似是一个“高大上”的事物，其实它与我们每一个大众的生活和工作都息息相关。不管你是大型企业的高层领导或普通职员，夫妻店或淘宝店店主，或是准备创业的年轻人，都能够利用基于数据分析的方法来帮助发现新的商机，提升你的管理，从而增强你的市场竞争力。比

如，利用历史的销售记录，可以识别出对你最有价值的顾客，从而通过精准营销等手段来实现最有效的客户关系管理。

要获得数据分析的能力，统计学是一把必须拥有的钥匙：数据分析最基础的工具非统计学莫属了。可以毫不夸张的说，统计学是应用最广泛、最具普适性的一门科学知识。它是人们把握不确定的世界里的基本规律，从而将客观的数据转化为有用的信息和知识的一把钥匙。只要你会一点统计分析的方法，也许会形成一套完全不同的看待问题的思维模式，从而帮助你更加聪慧地制定各种管理决策。

长久以来，统计学一直是大学课堂中的专利产品，好像与我们的大众没有丝毫的关系；因为它涉及到非常严谨的数理知识，对于没有系统学习微积分、概率论和线性代数的人来说只能是一种奢侈品。为了让大众都能充分享受统计分析带来的优势，日本帝京大学经济学系的小岛宽之教授结合自己长期的教学和科研心得，撰写了一本“可完全自学的统计学入门书”——《你一定爱读的极简统计学》。它面向数理基础薄弱的大众，在日本被誉为零基础的入门读物，已经热印18次，深受广大读者的喜爱。

一口气读完小岛宽之教授的专著，让我对大学教授这个职业的使命有了新的认识：从事国际前沿的学术研究固然是教授不可推卸的责任，向广大群众传播知识、让大众享受知识的红利同样重要。小岛宽之教授以最浅显的文字，采用深入浅出的方式，结合生活中的实际现象，向我们描绘了统计学的原理、方法与应用。最为难能可贵的是，学习本书几乎不用概率的知识，也完全不需要微积分和高等数学的基础，让零基础读者一看就懂，一学就会！只要读者有兴趣探索统计学的乐趣，就能通过自学的方式在很短的时间内理解统计学中标准差、置信区间、样本估计、样本比较等深奥的理论与方法，并能将该方法与身边的实际问题有效地联系起来。

我强烈推荐零基础的读者从这本书开始进入统计学和数据分析的神奇世界。也许你会发现，科学的管理和有效的决策并不是受过高等教育的人才能享受的专利；也许这本书会影响甚至改变你的一生。

2014年夏于清华园

前言

本书是一本关于统计学的书。

毫不夸张地说，本书是统计学的超级入门读物，面向：

- 统计学初学者；
- 想重新学习统计学的人；
- 几经挫折，（感到）至今仍未掌握统计学的人；
- 在统计学这门课程上，现在还处于落后状态的人。

它有其他教科书所没有的几个特征。这里只简单地列举几条，想更加深入了解的人，请阅读序章中的详细内容：

①“内容再削减就不能称之为统计学”，用最少的道具（工具）浅显易懂写成的“超级入门书”。

②几乎不用概率的知识。完全不用微积分和求和。只需要用到初中数学的知识（开方和一元一次不等式），所以即使不懂高等数学（即使全都忘记了）也没关系。

③每章附有填空式简单练习题，最适合自学。

④第1部分从最基础的知识开始，力求在最短的时间达到理解“检验”和“区间估计”等统计学最重要目标的目的。

⑤第2部分是对第1部分内容的深化。最高效地达到使用t分布进行小样本检验、区间估计这些可以称之为统计学最高级别应用的程度。只要理解这些，就能掌握统计学的要点。

⑥为了掌握标准差的意义，以简单的计算题和具体例子进行彻底解说。

⑦从统计学角度理解股票和信托投资等的投资风险，加强对金融商品的了解与把握。

那就让我们开始吧！

目录

第2部分 从观测数据推测其背后的广阔世界

序章　为了高效地、一步步理解“统计学”
——本书的立场

❶本书为什么由两部分构成

本书是统计学的入门书。我可以大胆断言，这是一本“内容再削减就不能称之为统计学”，最浅显易懂的“超级入门书”。

本书由两部分构成。第1部分从最基础的知识开始，力求在最短的时间达到理解“检验”和“区间估计”等统计学最重要的目标。

阅读第1部分，可以让我们在短时间内对“统计学要达到的目的以及如何实现”有整体上的了解。

现在正在为无论到哪学习统计学都“无法理解”它而抱头苦恼的人，或是无论阅读多少入门书却总是遇到相同难题的人，可以试着浏览这本书的第1部分。

这里一定有你想要理解却总是理解不了的内容。平日忙碌的读者，当你读到书中的某处时，你一定会感叹“原来统计学是这样的啊”，并认为本书物有所值。

第2部分是对第1部分内容的深化，解说关于母群体的推论统计方

法。第2部分的目标是最高效地达到使用t分布进行小样本检验、区间估计的程度。尽管只要理解这些，就能掌握统计学的要点，但很多学习者在此之前就已经备受挫折。

导致此情况最常见的原因就在于数据处理和概率这两部分。这两者几乎以同样的计算来定义，但原理该如何区分却极其难于理解——学习者大概就是因为这一点而陷入迷茫。

本书的第2部分，将包括数据处理与概率之间区别在内的，易使初学者陷入混乱的概念和枝节剪掉（在保证学术正确性的基础上），而选择统计的估计的本质结构，使读者能够直接地理解。也就是说，第2部分在某种意义上是对达到统计学重要目标的全力冲刺。

2 什么是统计学——描述统计和推论统计

大致而言，统计学由“描述统计”和“推论统计”两部分构成。

所谓描述统计，概括地说就是从取得的数据中抽取其特征的技术，起源可以说相当古老。比如，将人口调查作为一种数据来看的话，诞生了“摩西十诫”的摩西时代和罗马帝国时代等已经有了统计。汉朝时代的中国和大化革新时代的日本也有为了征税进行的人口调查和土地调查。

而描述统计学的确切起源在17世纪。

德国学者海尔曼·康令的《国势论》、英国军人约翰·格兰特的《关于死亡表的自然与政治的观察》以及威廉·配第的《政治算术》、爱德蒙·哈雷的《死亡率推算》等就是描述统计的先驱之作。在这些著作中，我们可以看出作者从有关出生率和死亡率的数据中明确地抽取出了特征，这正是站在描述统计学的立场上的研究方法。

此后，作为清晰抽取数据特征的工具，人们又开发出了频数分布表、直方图等图表方法，还有（各种）平均值、标准差等统计量方法。而现代人正在利用这些方法，对社会和经济状况进行把握，对气象和海洋等环境加以调查。

与此相对的推论统计，将统计学手法与概率理论相融合，对“无法整体把握的大的对象”或“还未发生而未来会发生的事情”进行推测。这是20世纪确立的方法论，从“部分推测整体”的意义上来说，即使称其为前所未有的全新科学也不为过。

就从我们身边来讲，选举速报可以算作是称为典型的推论统计的成果。在开票率仍在百分数阶段就可以进行“确定当选”的报道，这就是推论统计的功劳。此外，在全球变暖的预测、股票预测、金融商品和保险商品的定价等问题上，推论统计也是一种不可或缺的工具。

❸本书最重视标准差（*S.D.*）

本书第1部分的前半部分在解说描述统计时，选取了“标准差”为要点说明其意义。所谓标准差，是表示“数据在平均值周边分散程度”的统计量。笔者认为“标准差是统计学最重要的工具”，但很多统计学教科书只笼统地说明了其定义和计算方法。这使得学习者无法切身体会究竟“什么是标准差”。

而如果不能充分领悟标准差，在之后利用正态分布、卡方分布和t分布等展开推论统计时，就不能顺利地理解这些究竟是在做什么。这就是很多人学习统计学受挫的原因。

因此，本书从简单内容入手，从各个角度对标准差进行了解说，并自信这种解说在书中所占篇幅之大是其他教科书所无法企及的。说得具

体一些，本书不只是在单纯地提示定义，而是利用杂乱的公交车时刻表和冲浪者的比喻，还有股票指标等来使读者形象地理解其意义。而作为附加效果，读者还能理解在判断金融商品优良性上有重要作用的波动率和夏普比率。在21世纪高度发展的金融社会中，这些知识是非常有用的。

❹本书几乎不用“概率”

像序章中描述过的，为了将统计学应用于预测，必须在描述统计的方法上加上概率理论。描述统计学中学习过的平均值，在这里以随机变量中期望值的名称再次登场，而数据的标准差在随机变量中也以相同的标准差再次出现。虽然计算方法相同，但被当作不同的概念对待，就很容易使学习者产生混乱（实际上笔者在最初学习的时候也遇到过此问题）。

这种混乱在学习推论统计的进程中会变成大问题，最终导致学习者完全搞不清自己的学习内容是什么。

而之所以会出现这种混乱，原因在于统计和概率之间的微妙差异。统计是观测所得数据的集合，是“对于过去发生的事情的描述”。而概率，是“对于未来将发生的事情的描述”。所以，以“现在”为基准来看，两者意义完全不同。而若是从时间轴的往复来看，则可消除这种差异。

之所以这样说，是因为“未来发生的事情”在经过那一时点后，就变成“已经发生的数据”，而“过去发生的事情”追溯到那一时点之前，就成为“未来发生的事情”。对于这种微妙的既相同又有差异的统计和概率，使用平均值和标准差等相同的计算时，产生混乱也是在所难免的。而