



视觉新知口袋书系列

统计学入门 很简单

图解版

看得懂的极简统计学

[日] 涌井良幸 涌井贞美 著 刘楚姮 译

图解

使える統計学



中国工信出版集团



人民邮电出版社
POSTS & TELECOM PRESS



视觉新知口袋书系列

统计学入门

很简单

图解版

看得懂的极简统计学

[日] 涌井良幸 涌井贞美 刘楚姮

图解
使える統計学

人民邮电出版社

北京

图书在版编目(CIP)数据

统计学入门很简单：看得懂的极简统计学 / (日) 涌井良幸, (日) 涌井贞美著；刘楚姮译. — 北京：人民邮电出版社，2019.9
ISBN 978-7-115-51434-9

I. ①统… II. ①涌… ②涌… ③刘… III. ①统计学—基本知识 IV. ①C8

中国版本图书馆CIP数据核字(2019)第105676号

版权声明

ZUKAI TSUKAERU TOKEIGAKU

© 2015 Wakui Yoshiyuki, Wakui Sadami

First published in Japan in 2016 by KADOKAWA CORPORATION, Tokyo.

Simplified Chinese translation rights arranged with KADOKAWA CORPORATION, Tokyo through Beijing Kareka Consultation Center.

内 容 提 要

本书围绕搜集数据、分析数据以及得出统计结论这一主线，全面、系统地讲解了有关统计学的基础知识。其内容涉及统计原理、基本方法及发展应用，以及参数估计、假设检验、线性回归、实际应用等几个方面的有关知识。另外，本书还讲解 Excel 中几种常见的统计指标的计算方法，作为内容补充和完善。

本书文字简洁明晰，讲解简单易懂，注重结合实例介绍理论知识；不拘泥于繁杂的运算，深入浅出，从名词概念到原理方法，内容连贯，方便掌握，适合统计学初学者、对统计学感兴趣的读者以及想要学习运用统计知识的调研者使用。

◆ 著 [日] 涌井良幸 涌井贞美

译 刘楚姮

责任编辑 郭 媛

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：850×1100 1/32

印张：7.875

字数：197千字

2019年9月第1版

2019年9月北京第1次印刷

著作权合同登记号 图字：01-2017-6258号



定价：49.80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

译者序

放眼望去，我们生活的世界正在发生着深刻的变化，其变化速度已经不能再用日新月异来形容。甚至说，我们每一秒都处在信息爆炸之中。新闻、娱乐、广告、科技等信息像海浪一样将我们淹没。进入21世纪，人类的数据总量以指数增长的势态飞速膨胀，我们已经进入大数据时代。麦肯锡预言，数据将成为重要的生产因素。海量增长的数据为我们提供了无尽的可能性。例如在电商领域，运用人工智能对用户的购买喜好和浏览历史等数据进行深度解析，可以挖掘出用户需求，并进行精准营销。另外，通过数据识别把握市场的整体需求，可以进一步优化生产和运输，计算最短运输路径，安排最合适的生产计划，监控工艺和质量，进而提高供应链的效率。大数据还可以帮助我们处理社会问题，例如，利用人脸识别技术进行侦查破案，利用新媒体传播舆情实现社会舆论监督等。

在数据智能化的时代，统计学是我们探索世界的必备工具。世界越变越精彩，同时也越来越复杂。统计学是一种量化的手段。将真实世界中可能发生的事件赋予“概率”的含义，并基于概率计算出期望值。这样，我们可以在纷繁的选项中找出一条通向未来的路，这称为科学决策。同时，统计学又是一种把握问题的方法。将大量数据去粗取精，留下它们最想要讲的故事，这称为定量分析。

本书用通俗易懂的语言，图文并茂地讲解了统计学世界上最基本的理论知识，并且结合生动的案例解释了统计学在各个领域的应用，小到平常生活，大到治理国家，无一不与“统计学哲学”有关。本书极力想要向读者传达，统计学正是一门贴近现实的科学，并且它可以解释这个世界运行的基本规律。比如根据大数定律，人的幸运与不幸

虽然是一个概率问题，但是只要坚持不懈地尝试，幸运的天平终将会向努力的人倾斜。

这不仅是一本关于科学的书，更是一本富含哲理的书。对于译者来说，翻译的过程也是一个仔细研读的过程。虽然本书介绍的是统计学的基础，但是更着墨于基础知识的现实运用，非常具有启发性。当你遇到千头万绪无法归纳时，或是利益纠缠无法权衡时，不妨参考这本简单易懂的“统计学哲学”，或许可以为你点亮一盏灯。

刘楚姮

2019年2月10日于东京

序言

一百年以前，普通的人基本上没有接触到统计数据的机会。但是，在信息化高速发展的当今社会，我们正生活在信息的海洋里。因此，在现代社会里我们必须具备使用统计学处理信息的基本素养。通过学习统计学，大家可以掌握以下两项重要的能力。

第一，能够处理大量的数据。在信息化社会，面对工作和生活，我们必须能够正确处理手中的数据，并做出恰如其分的判断。

第二，能够正确理解数据分析的结果。当看到来自各种不同领域的统计数据时，大多数人倾向于认为所见即正确。然而，统计数据一定在某种程度上反映了数据创作者的意图，有时候甚至可能是人为捏造出来的。能正确理解数据传递出的信息是非常必要的。

本书是一本入门级别的介绍统计学基础知识的书，即便是只有初中数学水平的读者，也能轻松读懂。笔者真心地希望，通过阅读本书，大家能掌握统计学的基础知识，轻松应对现代生活。

最后，值本书发稿之际，向曾经悉心指导过笔者的角川中经出版社的若月孝之先生表达由衷的感激之情。

涌井良幸

涌井贞美

本书的使用方法

- 本书的难度相当于初中数学水平。学习本书内容，可以了解到现代统计学的基本思想。阅读本书时，不需要任何计算机方面的知识。不过，有时需要使用计算机解决书中复杂的计算问题（微软Excel软件的2007以上版本）。如果对本书涉及的计算机进行数据处理的内容有疑问的话，参考附录解说。
- 为了方便阅读，本书在计算时，数据都做了四舍五入处理。因此，求和时小数最后一位可能会对不上，求积的结果可能会有出入。数据的计算部分如果有错误还请见谅。
- 本书所使用的数据如果没有标注都是虚拟的。

目录

第 0 章 【序章】作为工具的统计学..... 1

为什么如今统计学备受推崇	2
统计学的分类	4
统计学的历史即人类的文明史	6
专栏 1 开放数据	8

第 1 章 【基础篇】数据的整理方法 9

数据种类繁多，主要分为定量数据和定性数据.....	10
列表整理使数据的特征显而易见.....	12
图表化使数据的特征一目了然.....	14
平均值是数据最重要的代表值.....	16
中位数 / 众数也是数据的代表值	18
方差 / 标准差是表征数据分散程度的重要指标	20
波动和方差的大小体现了原始数据的固有信息量	22
相关图使两个变量的关系可视化.....	24
相关系数使两个变量的关系量化.....	26
交叉表格使两个定性变量的关系可视化	28
专栏 2 多元分析	30

第 2 章 【基础篇】 概率 / 总体 / 样本 31

概率，即用 0~1 之间的数字来表示事件发生的可能性	32
随机变量，只有通过试验才能确认其数值	34
随机变量的概率分布是统计学重点	36
随机变量的平均值和方差的计算式	38
概率分布的百分点和 p 值	40
用于应对各种规范的标准化转换	42
总体及研究总体用的样本	44
统计学中经常使用的基本统计量	46
满足无偏估计量的条件	48
自由度取值不受限的变量个数	50
概率分布的基础是高斯发现的正态分布	52
集世界之美的中心极限定律	54
正态总体的样本均值的相关定理	56
正态分布经常用到的百分点	58
专栏 3 回归分析	60

第 3 章 【基础篇】 估计 / 检验的基本思想 61

从样本推测总体的统计学估计	62
区间估计的原理 1. 根据样本计算统计量的概率分布	64
区间估计的原理 2. 根据概率分布计算概率	66
区间估计的置信度 95% 及 99% 的意义	68
置信度和置信区间是此消彼长的关系	70
统计学检验的思想是简单易懂的	72
根据风险率（显著性水平）可知检验的正确程度	74

检验分为单侧检验和双侧检验两种类型	76
检验时无法避免的两类错误	78
专栏 4 因子分析	80

第 4 章 【基础篇】 回归分析 / 方差分析 / 贝叶斯统计学... 81

一元回归分析, 由 1 个变量预测其他变量	82
多元回归分析, 由 2 个以上的变量预测其他变量	84
决定系数决定回归分析的精度	86
自由度调整后的决定系数可以不受回归分析变量个数的影响	88
绝不是偶然的方差分析	90
组间和组内的变异是方差分析的决定性因素	92
方差分析的基础是使用 F 分布的检验	94
按部就班就可以完成方差分析	96
理解方差分析首先要理解偏差和波动的性质	98
双因素的方差分析 (无反复)	100
双因素的方差分析 (有反复)	102
方差分析的统计学术语	104
为什么现在贝叶斯统计学如此流行	106
4 个贝叶斯统计学的基本概念	108
基于条件概率的贝叶斯定理	110
贝叶斯定理的典型案例	112
专栏 5 主成分分析	114

第 5 章 【发展应用篇】数据的应用 115

调查 2 个商品价格之间的关联	116
-----------------------	-----

注意数据中的异常值.....	118
用大局观分析短期变化.....	120
调查区域差异带来的商品价格变化.....	122
画出商品价格的分布表.....	124
专栏 6 判别分析	126

第 6 章 【发展应用篇】 概率 / 总体 / 样本的应用 127

抽奖时首先要计算什么.....	128
从蒙特卡罗法中感受随机的威力.....	130
人的幸运和不幸与大数定律有关吗.....	132
喝醉的人的目的地是哪里	134
浅析中心极限定理.....	136
调查问卷的样本数量应该设为多少.....	138
专栏 7 聚类分析	140

第 7 章 【发展应用篇】估计 / 检验的应用 141

用极大似然估计法推导图钉针朝上的概率.....	142
从一个较大的样本中求主妇零用钱的平均值.....	144
从一个较小的样本中求主妇零用钱的平均值.....	146
已知方差求主妇零用钱的平均值.....	148
计算 30 ~ 40 岁男性的单身率.....	150
内阁的支持率的区间估计	152
调查商务人士零用钱的差异	154
调查吸烟率是否有增加.....	156
调查小学生数学的学习能力是否提高	158

调查新生儿的平均体重是否有减少	160
调查员工的睡眠时间是否有变化	162
调查东京和大阪的上班族的零用钱水平是否有差别	164
调查棒球选手的击球率是否有差异	166
调查葡萄酒实际容量的方差是否有变化	168
调查两个人的装袋工作的方差是否有差异	170
调查骰子是否有异常	172
调查某市男女出生比是否失衡	174
调查奶—茶和茶—奶是否可以靠味道分辨出不同	176
调查成人和儿童对西餐和日本料理的喜好是否不同	178
专栏 8 量化理论	180

第 8 章 【发展应用篇】

回归分析 / 方差分析 / 贝叶斯统计学的 应用..... 181

从住宅户数分析汽车保有量	182
从住宅户数和年收入分析汽车保有量	184
根据时间序列数据预测未来	186
展示新产品策略的效果	188
新教学方法是否有效	190
营养和睡眠对美容是否有效 (无反复)	192
补充	194
营养和睡眠对美容是否有效 (有反复)	196
补充	198
猜一下口袋里白球的个数	200
求解体检出现阳性结果时真的患病的概率	202
挑战蒙提霍尔问题	204

硬币出现正面的实际概率是多少	206
如何使用贝叶斯理论来筛选垃圾邮件	210
求解天气预报的概率	214

附录 A 使用 Excel 的求解方法..... 216

1. 用 Excel 求二项分布的 $100p\%$ 点和 p 值.....	217
2. 用 Excel 求正态分布的 $100p\%$ 点和 p 值.....	218
3. 用 Excel 求 t 分布的 $100p\%$ 点和 p 值.....	220
4. 用 Excel 求 F 分布的 $100p\%$ 点和 p 值	222
5. 用 Excel 求 χ^2 分布的 $100p\%$ 点和 p 值.....	223

附录 B 使用 Excel 分析案例 224

1. Excel 分析工具的安装方法 (Excel 2010, 2013).....	225
2. 第 4 章 [一元回归分析, 由 1 个变量预测其他变量] 的分析案例	226
3. 第 4 章 [多元回归分析, 由 2 个以上的变量预测其他变量] 的 分析案例	228
4. 第 4 章 [方差分析的基础是使用 F 分布的检验] 的分析案例	230
5. 第 4 章 [双因素的方差分析 (无反复)] 的分析案例.....	232
6. 第 4 章 [双因素的方差分析 (有反复)] 的分析案例	234
7. 第 8 章 [营养和睡眠对美容是否有效 (无反复)] 的分析案例.....	236
8. 第 8 章 [营养和睡眠对美容是否有效 (有反复)] 的分析案例.....	238

第0章

0

【序章】

作为工具的统计学

（统计学的必要性）

为什么如今统计学备受推崇

最近几年，统计学突然以惊人的速度流行起来。为什么它会受到社会的广泛关注呢？

最近几年，统计学突然以惊人的速度流行起来。为什么它会受到社会的广泛关注呢？

近年来，伴随着计算机性能的提升，我们的生活变得更加信息化。因此，现代社会也被称作信息化社会。在互联网上，每天都产生着大量的信息。真不愧是“大数据的时代”啊！身处这样的时代，具备统计分析的能力有如下两点好处。

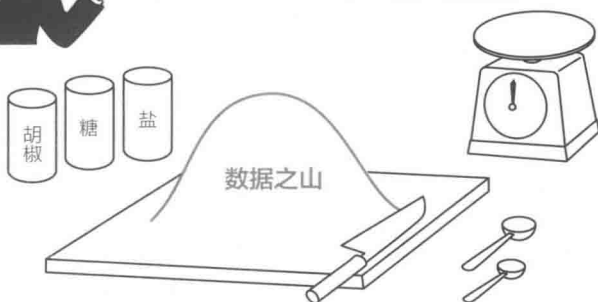
第一，能够处理大量的数据。在IT技术普及的当今社会里，如果不能掌握处理大量数据的能力，数据对我们来说不过是一座垃圾山。只要能掌握一点点的统计分析能力，我们就能把垃圾变为宝藏。在信息化社会，面对工作和生活，我们必须能够正确处理手中的数据，并做出恰如其分的判断。

第二，能够正确理解数据分析的结果。现在，各路媒体把来自不同领域的统计数据稍做加工，就大肆散布。然而，这些报道是否正确让人心存疑惑。很有可能，它们是被人为捏造出来的。如果我们稍微具备一点统计学的知识，就能识别出这些信息的真伪。

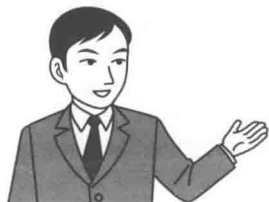
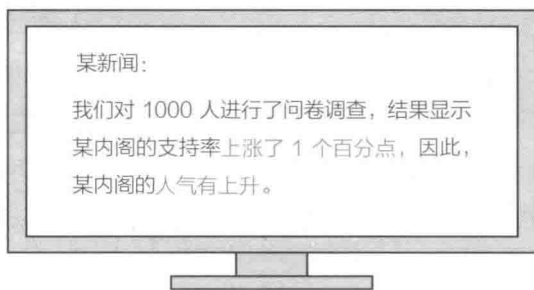
〔使用统计方法处理数据〕



把数据当作原材料，可以烹饪出各种各样的美食！



〔正确理解数据分析的结果〕



人气真的有上升吗？是不是误差呢？
本杰明·迪斯雷利有一句名言：“世界上有三种类型的谎言：谎言、该死的谎言以及统计数据。”

统计学的分类

统计学的目标是，理解数据中隐藏的本质信息。但是，根据达成其目标的方法不同，统计学可以分为不同的类型。

把整理后的数据做成表格或图表是描述统计学的常用方法。通过这样视觉化的处理，描述统计学将数据中隐藏的基本信息更直观地展示出来，便于读者理解。另外，描述统计学还运用平均值和方差这类简单的数值，直接体现数据的特征。

它能够如此简洁明了地展示数据的特征，因此成为展示研究成果、策划方案和商品说明等必不可少的重要方法。从前，我们总是在数据处理上耗费巨大精力，现在使用计算机就能轻松地完成了。

相对于描述统计学的是数理统计学，这是一种频繁使用到数学工具的统计学。推断统计学是其中一个分支，是一种用部分推测整体的统计学。推断统计学和描述统计学对概率的解释，是它们根本的分歧点。

很久以前，推断统计学还只是一种基于概率论（数据可以多次取样）的统计学，但是近年来，贝叶斯统计学（一种主张数据只能一次性获取的统计学）逐渐成为主流思潮。据说这种统计方法包含了经验性分析，相对容易理解。

此外，当存在大量变量时（例如，在公司销售、广告费用、员工数量等的情况下），这些变量的多元分析也是数理统计学的重要领域。