

后浪

The Lady Tasting Tea
How Statistics Revolutionized Science
in the Twentieth Century

女士品茶

统计学如何变革了科学和生活

[美] 戴维·萨尔斯伯格 (David Salsburg) 著
刘清山 译



了解统计学的人，运气都不会太差

科学松鼠会推荐统计学领域入门必读书

大师辈出的时代，
一场连环颠覆的思想盛宴

若想了解上帝在想什么，我们就必须学统计学，因为统计学在测量他的旨意。

——弗洛伦斯·南丁格尔

江西人民出版社
Jiangxi People's Publishing House
全国百佳出版社

女士品茶
统计学如何变革了科学和生活

[美]戴维·萨尔斯伯格 著
刘清山 译

版权信息

- 丨 书名：女士品茶：统计学如何变革了科学和生活
- 丨 作者：[美]戴维·萨尔斯伯格
- 丨 译者：刘清山
- 丨 版权：后浪出版咨询（北京）有限责任公司

谨以此书献给我42岁的妻子，亲爱的弗兰。我醉心于在职业生涯中收集统计革命先驱们的故事时，她一直在督促我将这些故事整理成一部不涉及数学符号的书籍。弗兰并没有接受过专业数学教育，她帮助我对本书进行了数次修改，指出我在哪些地方的解释不够清晰。这本书的成形，尤其是那些表述清晰的段落的成形，都要归功于她坚持不懈的努力。

你们不要回答调查问卷，
也不要询问尘世上的事情，
更不要顺从别人参与任何测试。
你们不要与统计学家为伍，
也不要研究社会科学。

——W. H. 奥登

为了理解上帝的思想，我们必须学习统计学，因为统计学
测量的是上帝的旨意。

——弗洛伦斯·南丁格尔

自序

科学进入19世纪，形成了一种坚定的哲学观念，这种观念被人们称为“按时钟前进的宇宙”。科学家相信，他们可以用少量数学公式（如牛顿运动定律和波义耳气体定律）描述现实，预测未来事件。这种预测只需要一组完整的公式和一组精度足够高的相关测量数据。普通民众花了40年时间才理解了这种科学观念。

这种文化差距的一个典型例子就是19世纪早期拿破仑皇帝（Emperor Napoléon）与皮埃尔·西蒙·拉普拉斯（Pierre Simon Laplace）之间的对话。拉普拉斯曾写就一部权威著作，描述了如何根据地球上的少量观测数据计算行星和彗星未来的位置。据说，拿破仑对拉普拉斯说：“我发现您在著作中没有提到上帝，拉普拉斯先生。”“我不需要这个假设。”拉普拉斯回答道。

对于并不存在上帝、没有神圣力量推动、按照时钟永远运行下去、一切未来事件由过去事件所决定的观念，许多人感到恐惧。在某种程度上，19世纪的浪漫主义运动就是对这种冷酷精准推理的回应。不过，19世纪40年代，这种新兴的科学理念获得了一个证据，令普通民众惊讶万分。科学家用牛顿数学定律预测了海王星的存在，而且人们在预测的位置发现了这颗行星。几乎所有反对“宇宙按时钟前进”的声音都消失了，这种哲学观念成了大众文化一个不可分割的组成部分。

不过，虽然拉普拉斯在公式中无须提到上帝，但他却需要所谓的“误差函数”。从地球上观测到的行星和彗星的数据与它们的预测位置并不完全吻合，拉普拉斯和同时代的科学家将其归结为观测误差。这种误差有时是由地球的大气扰动引起的，有时是由人为差错引起的。拉普拉斯把所有这些误差放在一个附加项里（误差函数），加入到他的数学公式中。这种误差函数吸收了所有误差成分，让拉普拉斯得到了预测天体真实位置的准确运动定律。人们相信，随着测量精度的提高，误差函数

最终会消失。有了用于解释观测值和预测值之间微小偏差的误差函数，决定论哲学掌控了19世纪早期的科学，人们相信一切事情的发生都是由宇宙初始条件和描述宇宙运动的数学公式事先决定的。

到了19世纪末，这种误差不但没有消失，反而变大了。随着测量精度的提高，人们发现了越来越多的误差。“按时钟前进的宇宙”开始松动。人们试图发现生物学定律和社会学定律的努力失败了。在更加成熟的科学领域如物理学和化学中，人们发现，牛顿和拉普拉斯使用过的定律只是一种粗略的估计。科学逐渐开始使用一种新的模式，即现实的统计模型。到了20世纪末，几乎所有学科都已经转移到了使用统计模型的阵营。

大众文化没有跟上这种科学革命的脚步。有些模糊的概念和表述（如“相关性”、“概率”和“风险”）成为了大众词汇，大多数人也知道了一些科学领域（如医学和经济学）上的不确定性，不过大多数普通人并没有意识到哲学观念已发生了深刻转变。这些统计模型到底是什么？它们是如何出现的？它们在现实生活中意味着什么？它们是对现实的准确描述吗？本书将尝试回答这些问题。同时，我们还会介绍与这场革命有关的一些重要人物。

在探讨这些问题时，我们必须区分三个数学概念：随机性、概率和统计。对大多数人来说，随机性只是不可预测性的同义词。《塔木德》中的一句格言可以体现这种流行观念：“不要寻找埋在地下的宝藏，因为宝藏埋藏的位置是随机的，从道理上说，我们无法寻找某种随机出现的东西。”不过，对现代科学家来说，随机性有许多不同的种类。概率分布的概念（本书第2章将会介绍）可以让我们对这种随机性加以限制，让我们获得预测未来随机事件的有限能力。因此，对现代科学家来说，随机事件并不是没有规律、出乎意料、无法预测的，它们拥有一个可以进行数学描述的结构。

概率是一个现代词汇，但它表达的概念非常古老。亚里士多德（Aristotle）曾说：“概率使那些不太可能发生的事情得以发生。”最初，这个词语表达了人们对可能发生的事件的感受。17、18世纪，包括伯努利家族（the Bernoullis）两代人、费马（Fermat）、棣莫弗（de Moivre）、帕斯卡（Pascal）在内的一批数学家共同建立了概率的数学理论，这个过程始于概率游戏。他们提出了一些非常复杂的方法，用于计算等概率事件。棣莫弗成功地将微积分方法应用到了这些计算中，伯努利家族则发现了一些非常基本的定律，叫做“大数定律”。到了19世纪末，概率

数学主要由复杂的技巧组成，但是缺乏坚实的理论基础。

虽然概率理论并不完善，但它对统计分布概念的提出居功至伟。在我们考虑一个具体科学问题时，往往需要用到统计分布。例如，1971年，哈佛大学公共卫生学院在知名医学期刊《柳叶刀》上发布了一份研究报告，研究喝咖啡与下尿道癌是否存在相关性。该报告研究了一些病人，其中有些人患有下尿道癌，有些人患有其他疾病。报告作者还收集了这些病人的其他数据，如年龄、性别、家族患癌史。不是每个喝咖啡的人都得了尿道癌，也并非每个尿道癌患者都喝咖啡，所以有一些事件与他们的假设相矛盾。不过，25%的尿道癌患者经常每天至少喝四杯咖啡，而只有10%的非尿道癌患者喝咖啡达到这样的数量。看起来，似乎存在支持这种假设的某种证据。

报告作者收集的数据形成了一种统计分布。利用概率数学工具，他们为这种分布构造了一个理论公式，叫做“概率分布函数”，简称分布函数，用于研究这个问题。这个函数类似于拉普拉斯的误差函数，不过更为复杂。分布函数理论的构造用到了概率理论，他们用该函数描述从同一人群随机获取的未来数据的期望值。

本书并不是一本介绍概率和概率理论这些抽象数学概念的书。本书介绍的是一些概率定律在科学问题、统计分布领域及分布函数方面的应用。概率理论本身不足以描述统计方法，有时科学上的统计方法还会违反某些概率定律。读者将会发现，概率这个概念在本书各个章节中时隐时现，在需要的时候会提及，不需要的时候则会忽略。

现实的统计模型属于数学模型，只有通过数学公式和数学符号才能让人充分理解。我不想把这本书写得特别复杂，我只是通过20世纪科学统计革命中的一些相关人物（许多人仍然健在）来描述这场革命。我只是介绍了他们的工作，以便让读者感受到他们的个人发现是如何融入整个革命进程中的。

本书无法让读者学到足够多、可用于科学数据分析的统计方法。这需要数年研究生课程的学习。不过，我希望读者阅读本书之后能够对科学统计思想体现出的哲学基础的深刻转变获得一些理解。那么，对于一个对数学知之甚少的人来说，从哪里开始了解这场科学革命呢？我觉得品茶的女士是一个不错的选择……

目录

自序

第1章 品茶的女士

科学的合作本质

实验设计

第2章 偏斜分布

高尔顿的生物统计实验室

相关与回归

分布与参数

《生物统计》的计划

第3章 亲爱的戈塞特先生

“学生”的诞生

“学生”的t检验

第4章 堆积如山的记事本

费希尔与卡尔·皮尔逊的对立

“法西斯主义者”费希尔

《研究工作者的统计方法》

洛桑与农业实验

第5章 《收成变动研究》

《收成变动研究一》

高尔顿均值回归的推广

随机对照实验

费希尔的方差分析

自由度

《收成变动研究三》

第6章 “百年一遇的洪水”

极值分布

标志性作品

第7章 费希尔的胜利

费希尔和皮尔逊的统计观点

费希尔的最大似然方法

迭代算法

第8章 致死剂量

概率单位分析

布里斯的后期工作

第9章 钟形曲线

欢迎访问：电子书学习和下载网站 (<https://www.shgis.com>)

文档名称：《女士品茶：统计学如何变革了科学和生活》[美]戴维·萨尔斯伯格 著.pdf

请登录 <https://shgis.com/post/2241.html> 下载完整文档。

手机端请扫码查看：

