

C.Radhakrishna Rao

统计与真理

怎样运用偶然性

[美] C.R. 劳 / 著

 科学出版社
www.sciencep.com

统计与真理

—— 怎样运用偶然性

[美] C.R. 劳 著

科学出版社

北京

内 容 简 介

本书是当代国际最著名的统计学家之一 C. R. 劳的一部统计学哲学论著,也是他毕生统计学术思想的总结,同时还是一本通俗的关于统计学原理的普及教科书。

书中,作者从哲学的角度论述了统计学原理,通过实例,不仅证明了统计学是一门最严格、最合理的认识论和方法学,还深刻地揭示了现代统计学发展的过程,特别是那些很深刻的理论是如何从一些非常简单实际的问题中发展起来的。本书前 5 章讲述了统计学从最初收集、汇编数据为行政管理服务,发展成为有一整套原理和研究方法的独立学科的历史,第 6 章谈了普通公众对统计学的理解,强调了从数字中学习有助于成为有效率的公民。本书最引人注目的特点是,书中提到的所有科学的学科调查与决策和统计之间的关联是由一系列实例来说明的。本书使用非专业语言通俗地阐述了统计学的基本概念和方法,适合大众读者。

图书在版编目(CIP)数据

统计与真理:怎样运用偶然性/(美)C. R. 劳著. —北京:科学出版社, 2004

ISBN 7-03-012222-4

I. 统… II. 劳… III. 统计学—通俗读物 IV. C8-49

中国版本图书馆 CIP 数据核字(2003)第 088019 号

责任编辑:吕 虹 陈玉琢/责任校对:宋玲玲

责任印制:钱玉芬/封面设计:黄华斌

科学出版社 出版

北京东黄城根北街16号

邮政编码:100717

<http://www.sciencep.com>

新蕾印刷厂 印刷

科学出版社发行 各地新华书店经销

*

2004年7月第 一 版 开本:B5(720×1000)

2004年7月第一次印刷 印张:9 1/4 插页:1

印数:1—5 000 字数:162 000

定价: 22.00 元

(如有印装质量问题,我社负责调换〈环伟〉)

本书谨献给
引导我探求知识的
母亲

A. Laxmikanthamma

在我年少时,母亲每天早上四点起床,为我点上油灯,
使我能在安静的早晨精力充沛地用功读书

知识是我们已知的
也是我们未知的
基于已有知识之上
我们去发现未知的

由此,知识得到扩充

我们获得的知识越多
未知的知识就会更多

因而,知识的扩充永无止境

* * *

在终极的分析中,一切知识都是历史
在抽象的意义下,一切科学都是数学
在理性的基础上,所有的判断都是统计学

中文版出版说明

我很高兴看到《统计与真理——怎样运用偶然性》一书走进中国。我要感谢李竹渝博士、石坚博士和白志东教授，因为我的书稿是英文的，为了把本书奉献给中国读者，他们对书稿的翻译做了大量的工作。

如何建立新知识？这个问题取决于我们对知识概念的认知。以一个哲学家的观点，知识存在于（真实的或确定的）谬误之中，推理是获得这样的知识的工具。而从一个科学家的观点来看，一切知识都不是绝对正确的。通过任何方式所得到的一个科学理论知识，如果能引导出可接受限度内的预示，就能获得认可。一个新的理论如果能提供更好的预示，就将取代已经存在的科学理论。而从统计学的角度来看，从经验或实验中获取的知识是不确定的，但在实际生活中，不管这些已有的知识如何贫乏，我们不得不以此做出决策。统计学关注的是如何探知由观察数据获取的知识中的不确定性的量度，以及如何明确在最小损失下的最优决策。

《统计与真理——怎样运用偶然性》讨论的问题是：如何设计实验以便提供所要求的信息，如何从实验结果中获取一切有效信息，以及如何在实际中应用这些信息。本书将通过大量的事例来讨论这样一个过程中的哲理、逻辑和实践的问题，以及如何避免对统计方法的误用或对统计学的误解的问题。

人类一切努力的最终目的是寻求真理，而在严格意义下的真理是不可得到的，替代的是要寻求可接受的知识。严格地讲，知识不是真理，但它应最好地运用真理。我们应该记住中国大哲学家孔子所说的：知之为知之，不知为不知，是知也。最后，我要感谢科学出版社在本书的出版过程中的合作和努力。

C. R. 劳

宾夕法尼亚州州立大学大学城

2003年7月28日

不确定性知识

+

所含不确定性量度的知识

=

可用的知识

序

《统计与真理——怎样运用偶然性》一书的英文版,是在1987年纪念拉曼纽扬的百年诞辰活动期间、我所作的三次关于统计学历史与发展的演讲的基础上完成的.演讲中所涉及的每一个主题都更详细地重新写过,各自独立成章.现在的中文版在许多方面与英文版有明显的不同.

英文版中的第一、第二和第三讲的内容现在完全被重新组织,并扩展成为系列的5个章节,回顾了统计学从最初收集、汇编数据为行政管理服务,发展成为有一整套原理和研究方法的独立学科的历史.书中提到的所有科学学科的调查与决策和统计学之间的关联由一系列实例来说明.最后一章,即第6章是新增加的,谈及普通公众对统计学的理解,也是大家感兴趣的话题.

第1章所涉及的随机性、混沌与偶然性这些概念在调查和解释自然现象中扮演着重要的角色.强调了随机数在秘密通讯、产生无偏差信息以及在解决复杂计算中的重要性.也谈到了与艺术和科学中的创造性有关的一些想法.第2章介绍了在新知识建立中常用到的归纳法与演绎法.同时也说明如何量度不确定性使人们能获得最佳决策.

统计学思想远古即存,但作为一门学科历史却不长.第3章和第4章从原始人用刻痕记数来记录他的牲畜的数量开始,到从数字或者已知数据中抽取信息,成为在不确定性条件下做出推断的一种强有力的逻辑工具方面,来讲述统计学发展的历史.也强调了避免伪造数据、污染数据或任意对数据进行编纂的重要性,同时介绍了如何检测数据中存在这一类问题的一些方法.第5章用来讲述统计学的无处不在,无论在解开自然奥秘的科学调查中,或是要在日常生活中做出最佳决策,或者要解决法庭争端时,统计学都是一种探求真理的必不可少的工具.

我们都生活在信息时代,大多数的信息都是以量化的形式传播的.例如:今年的犯罪率与前一年相比下降了10%;明天有30%的可能要下雨;股票市场的道·琼斯指数价格增加了50点;世界上每4个新生婴儿中有一个是中国人;赞成总统的外交政策的人数占总人数的57%.这个估计的误差不会超过4个百分点;如果你坚持独身,你的寿命要减少8年.所有这些数字对一般公众来说到底意味着什么呢?这些数字里面包含什么样的信息会有助于个人做出正确决策去改进提高他们的生活质量呢?强调公众对统计学理解的需要是在本版新加的第6章里所作的一个尝试.能从数字中学习有助于成为有效率的公民,正如韦尔斯(H. G. Wells)所强调的:

统计思维总有一天会像读与写一样成为一个有效率公民的必备能力。

1987年,在每次演讲的开始,我都要提到拉曼纽扬的生活和工作.我将所有这些传记性的细节作为一个与拉曼纽扬生平有关的文献放在了本书末的附录里。

C. R. 劳

译者的话

自1992年11月我国国家技术监督局颁布的GB/T14745-92《学科分类与代码》中,将统计学与数学、经济学等学科并列上升为一级学科,把包括原属社会科学领域和自然科学领域的各种统计学归并为一门统计学以来,统计学学科的发展和统计学教育就以一个新的面貌在我国出现了.在我国,由于长期受原苏联统计学教学思想的影响,理科将统计学视为一门数学专业,文科方面仅注重宏观统计描述,没有充分重视统计学科学性的研究,没有充分重视统计分析和数据计算分析对解决实际问题的潜在功能,使我国统计专业的发展一方面长期偏向数学理论,忽视统计分析本身与多科学交叉发展的内涵;另一方面带有较强的计划经济下的定性分析角色,失去了以数据来解释现象的统计学的本质.近十几年以来,我国逐步大力开展统计分析的实际应用,开始注重统计学与其他科学交叉发展的理论与应用研究.但与国际水平相比,在利用现代统计学的理论、方法和计算功能来解决自然科学、社会科学中的实际问题方面,我们还有相当的距离;交叉学科的发展,如生物统计、金融统计、经济统计等方面,我们还没有形成较强的研究实力.一般公众对统计学的认识还有待进一步提高.

本书是统计学界最知名的权威之一C. R. 劳的著作,是他毕生经验的总结,既是一本高深的统计学哲理的专著,又是一本通俗的统计学原理的普及教科书.(本书跋的作者白志东先生已对C. R. 劳先生毕生对统计学的贡献及本书的内容作了较详细的介绍,这里不再赘述.)自英文版问世以来,在世界各地广泛流传,并先后出版了日文、西班牙文、波兰文、德文和中文繁体字版等.我们和本书作者C. R. 劳先生商量出版中文简体字版,他欣然同意,并委托我们做中文简体字版的翻译和出版工作.相信本书中文简体字版的出版将对我国学生学习统计学知识有一定的帮助.

在准备中文简体字版期间,欣喜得知C. R. 劳先生荣获2002年度美国总统科学奖,并于2002年6月12日在白宫接受了布什总统的颁奖,表彰他在“统计学理论的建立,多元统计分析方法及其应用方面所做的开拓性贡献,其丰富了物理学、生物学、数学、经济学和工程学的发展”.我们谨以中文简体字版的正式出版作为我们对先生获奖的衷心祝贺,并庆贺先生83岁诞辰.

戴维·柯克斯爵士(David Cox Sir)评论本书原文时曾谈到“书中论题涉及了从创造性本质这样一些一般哲学概念到专业统计学原理,是一本阐述统计学论点本质的力作”.尽管我们力求准确把握原文内容,但由于学力有限,专业和文字能力则

有未逮,译文中定会存在词不达意甚至误译的地方,诚恳专家、学者和广大读者的批评指正.另外,我们在翻译的过程中,专业人名主要参考了《英汉数学词汇》(科学出版社),《英汉统计学词汇》(中国统计出版社),《英俄汉数学词汇》(广东科技出版社).个别找不到译法的,没有译出.

我们感谢 C. R. 劳先生和科学出版社给我们这个机会,感谢白志东先生专门为本简体字版作跋,同时,感谢鲁万波、李方文在本书文字编辑方面给予的大力支持.

译 者

川大花园,成都

2002年12月

就像房屋是由石头堆砌的一样，科学是由事实构成的。
但如同一堆石头并不是一栋房子，仅仅是事实的收集，也并不
成为一门科学。

J. H. Poincare

对统计学的一知半解常常造成不必要的上当受骗
对统计学的一概排斥往往造成不必要的愚昧无知

目 录

第 1 章 不确定性、随机性与新知识的创立	1
1.1 不确定性及其度量化.....	1
1.2 随机性与随机数.....	2
1.3 从决定论到无序中的有序.....	13
1.4 随机性与创造性.....	15
参考文献	18
附:讨论.....	19
A.1 偶然性和混沌.....	19
A.2 创造性	20
A.3 偶然性和必然性	24
A.4 模糊性	26
A.5 π 的小数点后的位数是随机的吗?	27
第 2 章 不确定性的驾驭——统计学的发展	29
2.1 早期历史:作为数据的统计学	29
2.2 不确定性的驾驭.....	33
2.3 统计学的未来.....	40
第 3 章 数据分析的原理和策略——数据的交叉检验	43
3.1 数据分析的发展历史.....	43
3.2 数据的交叉检验.....	47
3.3 媒介分析.....	58
3.4 推断数据分析与结束语.....	59
参考文献	60
本章没有引用的附加参考文献	62
第 4 章 加权分布——有偏数据	63
4.1 设定.....	63
4.2 截断分布.....	64
4.3 加权分布.....	66
4.4 随机比率抽样法(p. p. s. 抽样法).....	67
4.5 加权二项分布:经验定理	68
4.6 酗酒,家庭人数与出生顺序	74

4.7 等待时间悖论·····	77
4.8 损伤模型·····	78
参考文献·····	79
第5章 统计学——探求真理必不可少的工具 ·····	81
5.1 统计与真理·····	81
5.2 某些实例·····	86
5.2.1 莎士比亚的新诗：一曲统计学的赞歌·····	86
5.2.2 有争议的作者权：联邦主义者论文集·····	88
5.2.3 卡尔特亚与《印度经典》·····	89
5.2.4 出版年月·····	89
5.2.5 柏拉图著作的系统排列·····	89
5.2.6 原稿的鉴定·····	90
5.2.7 语言树·····	90
5.2.8 地质年代的尺度·····	91
5.2.9 鳗鱼的公共繁殖场所·····	92
5.2.10 人所具有的特点是遗传的吗？·····	92
5.2.11 左撇子的重要性·····	93
5.2.12 日内循环·····	95
5.2.13 辨明生父·····	96
5.2.14 统计学中的盐·····	96
5.2.15 血液检查中的经济学·····	97
5.2.16 为增加粮食生产而建设机械工厂·····	98
5.2.17 小数位数字的遗失·····	99
5.2.18 Rh(Rhesus)因子：科学的调查研究·····	100
5.2.19 家庭人口、出生顺序和智商 I. Q.·····	101
参考文献·····	102
第6章 统计学的公众理解——从数字开始学习 ·····	104
6.1 大众的科学·····	104
6.2 数据、信息和知识·····	105
6.3 信息革命与统计学的理解·····	107
6.4 令人悲哀的数字·····	109
6.5 天气预报·····	111
6.6 社会舆论调查·····	112
6.7 迷信和心理作用·····	113
6.8 统计学与法律·····	114

6.9 超灵感与惊人的巧合	116
6.10 普及统计能力.....	117
6.11 统计学, 一门关键的技术	118
参考文献.....	118
附录 拉曼纽扬(S. Ramanujan)——一位罕见的天才	119
索引.....	123
跋.....	128

第 1 章 不确定性、随机性与新知识的创立

让混沌涌来吧！

让云彩形成一片沼泽！

我等待着成形。

罗伯特·弗罗斯特(Robert Frost)

1.1 不确定性及其度量化

不确定性与随机性的概念已经困扰人类很长一段时间了。在我们生活的物质世界和社会环境中，我们无时无刻不面对不确定性，遭受各种自然灾害，忍受着大自然的不确定性，正像歌德所想像的那样，事物是具有不确定性的：

伟大的、内在的永恒不变的法则能给我们指出使我们不再徘徊的路吗？

或者是像近三个世纪以来，也可以说是从古至今最伟大的物理学家爱因斯坦所相信的那样：

上帝决不会和宇宙赌博。

某些神学家认为：因为上帝决定世间万物，对上帝来说没有什么是随机性的。但也有人断言：即使是上帝，也被某些随机现象所左右。弗朗斯(A. France)在他所著的《伊比古鲁的乐园》(The Garden of Epicurus)一书中写到：

所谓随机性，恐怕是当上帝不愿显示其真实身份时所用的托词而已。

从亚里士多德时代开始，哲学家们就已经认识到随机性在生活中的作用，他们把随机性看做破坏秩序规律和超越人们理解能力范围的东西，但没有认识到有可能去研究随机性，或者是去测量不确定性。印度的哲学家们信奉古印度的因果报应学说，认为没有必要去研究随机性，因为按其严格的因果关系教规解释：一个人的命运，是由人的前世的行动所决定的。

所有人类的活动都是基于某种预示的，如上大学，找工作，结婚或投资。既然未来是不可预测的，不管人们掌握多少信息，都不可能存在能做出正确决策的系统方法。做出决策时，为了避免不确定情况和防止产生错误，人们依赖于像占星术