STIC 科学与管理译丛

信息专业人员常用统计方法

用实用、轻松的方式来理解、使用和阐释统计学

Statistical Methods for the Information Professional

A Practical, Painless Approach to Understanding, Using, and Interpreting Statistics

Liwen Vaughan(邱黎雯) 著

张爱霞 吴雯娜 等 译
李 燕 潘晓蓉

张新民 等 校

科学技术文献出版社

科学与管理译丛

信息专业人员常用统计方法

用实用、轻松的方式来理解、使用和阐释统计学

Liwen Vaughan(邱黎雯) 著
张爱霞 吴雯娜 等 译
李 燕 潘晓蓉
张新民等 校

科学技术文献出版社
Scientific and Technical Documents Publishing House
北 京

译者简介

张爱霞　北京大学信息管理系在读博士，中国科学技术信息研究所馆员

吴雯娜　硕士，中国科学技术信息研究所副研究馆员

李　燕　硕士，中国科学技术信息研究所馆员

潘晓蓉　硕士，中国科学技术信息研究所馆员

尹盛鑫　双学士，中国科学技术信息研究所工程师

杨代庆　硕士，中国科学技术信息研究所馆员

刘春燕　硕士，中国科学技术信息研究所馆员

张新民　博士，中国科学技术信息研究所副研究员，硕士研究生导师

译　序

在这个“用数据说话”的年代，统计学基本知识与分析技能对于任何行当都非常重要。从正面说，掌握了统计学方法，就有利于收集适当的数据，开展深入的分析，得出有理有据毫不牵强的结论，从而说服别人；从反面说，不掌握统计学原理，就很容易受骗上当。曾任19世纪英国首相、同时又是小说家的本杰明·狄斯雷利(1804—1881)曾经幽默地说过，“世界上有三种类型的谎言：谎话、弥天大谎、还有统计数字”，就生动地说明了不懂统计学和误用统计学的危险。

在迈向知识社会的进程中，图书情报事业的重要性受到空前的重视，从而对图书情报人员的业务素质提出了更高的新要求，掌握基本统计学知识与技能便属于新要求之一。而我们面临的现状是，无论在中国还是在国外，图书情报学从业人员中能完善地掌握统计学方法的不多，一知半解的倒不算太少。英国诗人蒲伯(1688—1744)说，“一知半解可能是很危险的”，因为，拥有一知半解的人，往往自以为已经懂得不少了，容易率尔操觚，贻笑方家。

曾经在上海科技情报研究所工作、目前在加拿大西安大略大学任教的邱黎雯教授对于此类问题有深刻的观察与体会。为了帮助图书情报人员更有效地、更有针对性地学习和掌握统计学知识，她发奋

努力编写了这本教材。教材推出后，在国际图书情报学界获得好评，因为过去从来没有专门为图书情报人员编写的类似教材。我们组织力量将此书译出，也是为了填补中文图书市场上的这一空白，更好地满足中国图书情报专业的学生和从业人员的需求。在此，我们向本书作者邱黎雯教授表示深深的感谢和敬意。

在全球化的世界上，保持多样化显得更加重要。翻译工作是在多样性的文化之间架设桥梁的伟大工程。笔者曾写过一首打油词“西江月·译事咏”，上半阕说的是：“文化相隔如岛，翻译可作津梁。互通有无较短长，委实功德无量”，就反映了我们对翻译工作的认识。多年来，中国科学技术信息研究所一直注重科技翻译工作，推出了不少有影响的译著。遗憾的是，翻译劳动的价值迄今未受到应有的重视，翻译稿酬之低暂且不说，翻译成果在评职称、报奖时要么不被承认，要么大打折扣。但是，只要社会需要翻译作品，我们就会一如既往、义无反顾地做下去。

中国科学技术信息研究所总工程师

武夷山

2007 年 11 月 25 日

译者的话

本书译自 Information Today 公司出版的邱黎雯教授所著的《Statistic Method for the Information Professional》。邱黎雯系加拿大西安大略大学教授，长期从事研究方法和统计学、网络搜索引擎、网络数据挖掘、信息检索以及情报计量学等方面的教学和科研工作。本书是作者教学和科研实践经验的总结，深入浅出地讲述统计学方法在信息科学领域中的应用，侧重统计方法的基本逻辑，而不是繁杂的数学推理；强调的是何时运用和怎样运用统计学方法以及如何理解和解释统计分析结果。对于我们而言，无论是做统计分析，还是理解在阅读文献中遇到的统计学问题，这本书都是非常有用的。

本书由张爱霞联合多人共同协作完成。各章的翻译分工如下：第 1、2 章和索引部分张爱霞；第 3 章刘春燕；第 4、5 章及前言部分吴雯娜；第 6、7 章李燕；第 8、9 章及第 12 章尹盛鑫；第 10、11 章杨代庆；第 13 章及附录部分潘晓蓉。为确保翻译质量，我们对译文进行了多次修改和校对。张新民、庞景安承担了译稿的审校工作，最后由张新民进行了终校终审。

在翻译中，为便于读者阅读查找，原著索引在译著中以中英对照的方式出现，此外，对于索引中出现的重要词汇，其在译著正文中首次或在重要位置出现时都标出了相应的英文，以便于理解。

在此，我们想对大家表示诚挚的谢意：感谢中国科学技术信息研究所总工武夷山研究员，是他推荐此书，并在百忙之余拨冗为本书作序，使得本书大为增色；感谢中国科学技术信息研究所信息资源中心主任沈玉兰研究馆员，她对本书的翻译和出版工作给予了极大的帮助和支持；感谢林明、陶冶、李建丽、蔡小英、刘智、蔡晓燕等人在本书校对中给予的支持。

译文虽然经过多次校对和修改，但由于译者水平有限，加上时间仓促，文中难免有不妥之处，我们真诚地希望同行和读者不吝赐教。

译者

2007 年 11 月 1 日

关于作者

邱黎雯(Liwen Vaughan)博士于1991年在加拿大西安大略大学获得图书情报学博士学位。她对统计分析有着浓厚兴趣,同时,她又能以清晰和可理解的方式对复杂的概念进行解释,这使得她成为一个成功的教育家。她有10年以上面向不同专业和不同学位学生讲授统计学的教学经验,涉及专业包括图书情报学、新闻学和企业管理等,授课学生包括学士、硕士和博士。邱黎雯博士在情报学研究中有丰富的统计分析经验。她的研究成果发表在很多情报学期刊中,在这些成果中,她运用了许多统计方法。她的关于用Markov模型分析超文本信息的论文,发表在JASIS上,曾经吸引了全世界的博士生的关注。最近,她成功地利用LISREL模型(一种新的应用到情报学中的高级统计方法)定量地研究了信息对企业成功的影响。

致　谢

感谢我的丈夫 David Vaughan，他在本书编写过程中给予了我精神上和技术上的支持。他精心编辑了全书，不辞辛劳地把书中所有的数字和图表转换为要求的出版格式；更重要的是，他非常支持我的见解和想法。没有他的帮助，本书不可能完成。我的儿子 Ulysses，在我着手撰写此书时尚在腹中，到我完成这本书时他已经在蹒跚学步了。确切地说，他是伴随着我的写作一点点长大的。他一直都是我的灵感来源。写这本书常常使我顾不上照顾家庭，对于 David 和 Ulysses 我有着深深的负疚。谨将此书献给他们。

Liwen Vaughan(邱黎雯)

前　言

不仅仅是又一本统计学书

统计学是一种最有用和最有力的数据分析工具，同时也是一个最令人生畏和最令人生厌的研究主题，这种现象很值得注意。之所以令人畏惧，是因为统计学著作通常充斥着大量公式、数学术语以及连篇累牍的推导和演算。本书与其他众多的同类书籍有所不同，不同之处在于其核心是理解统计学的基本逻辑，而不是错综复杂的数学难题；它强调的重点是统计学的含义、应用场合、应用方法和对结果的解释。本书反映了我从十年的统计学教学中提炼出来的非常成功的方法。这种方法具有以下三个特点：

第一，我采用逻辑推理而不是数学推导的方式去解释统计学概念和统计检验。统计学虽然是数学的一个分支，但很多统计检验的内在逻辑非常直接，不需要具备高等数学知识背景就可以理解。经常在统计学著作中见到的大量篇幅的公式和数学解释，往往掩盖了这种内在逻辑的简单性。所以，在本书中，你不会看到太多的公式。你看到的少数几个公式也会有详细的解释，而且，你也可以忽略这些公式，这并不会妨碍你理解正在讨论的基本原理。

第二，我采用信息科学研究中的实例来说明所涉及到的每个统计检验的过程。我先从构造假设开始，然后用计算机软件分析数据，整理得出的分析结果，最后得出结论。读者可以看到统计学方法应用的完整过程，而不会陷入到技术细节中去。这种强调统计学方法的实际应用的做法可以使读者明白统计学是一个多么有力和适用的工具，而不是与他们无关的晦涩的数学分支。

第三，我强调要利用计算机软件来进行演算和完成其他繁琐的数学工作。很多统计学著作花费大量时间去做公式推导和演算。事实上，现在几乎不会有人用手工方式来完成这些演算。现在已有各种各样的计算机软件包来完成繁

重的数学工作，因此我们就可以将重点放在理解结果的含义上。有鉴于此，我用了一章的篇幅来论述有关使用计算机软件进行统计分析的问题。

我采用这种方法来教我的学生学习统计学。他们来自各个学科，从本科生到博士，各种层次的都有。我的大部分学生几乎没有什么数学基础，许多学生刚接触这门课程时都感到畏惧和紧张。然而，令他们惊喜的是，统计学竟然如此使人轻松，弄懂一个对他们来说曾经是神秘和陌生的学科如此使人兴奋。事实上，认为统计学很复杂，非数学专业人员难以企及，是一种误解。我相信，一旦你读了这本书，就会赞同我的看法。

本书的读者对象

本书适用于信息专业人员，包括学术研究人员和实际从业人员。他们将会发现，这本书在理解统计学的概念和技巧上非常有用，这些概念和技巧在信息科学中的应用越来越广泛，而且在文献中也大量出现。作为信息专业人员，即使你不亲自去做统计分析，你至少也会使用别人做出的统计结果。本书的目的就是为你提供一种帮助你理解和应用统计学方法的策略。即使你不是信息专业人员，也可以从此书中受益。归根结底，统计学的基本理论和原理对于任何领域都是一样的。

如何使用本书

本书涵盖了信息科学研究中常用的基本统计学方法。这些方法包括：描述统计学、t 检验、χ^2 检验、方差分析、相关、回归和基本非参数统计检验。书中也将会介绍一些中级和高级的统计方法，比如多重回归分析和 LISREL。

如果你已经熟悉一般的统计分析方法，使用本书是为了重温某一主题，那你可直接跳到你感兴趣的章节。第 12 章中还特别给出了一个总图，你可根据实际情况，依据此图来决定要复习哪种检验方法，然后再去相应的章节阅读。

如果你是个统计学初学者，或者你过去曾经学过，但需要再复习一下以前学过的一些主要概念，我建议你依次读完前 6 章，然后再看其他章节。第 1 章讨论如何区分不同类型的数据，这是开始进行数据分析时最先遇到的问题。第 2 章涉及使用计算机软件进行统计分析的一般性问题，包括如何将数据组织到计算机文件中，以及怎样处理缺失数据。第 3 章和第 4 章分别讨论了图表和描

述统计学。第5章介绍了推断统计学的基本概念。因为本章是后面讨论的所有统计检验的基础，所以在转入后面的章节之前应该先阅读此章。第6章讨论了各种抽样方法。

前6章的性质决定了读者需要按章节顺序依次阅读，但从第7至11章就不必依次阅读了，因为每章介绍的是不同的统计检验方法。所以你可以直接转入你特别感兴趣的一章。第12章对第7至11章中所涉及到的推断统计学检验进行了总结，其中，有一个“路线图”总图，用以引导你根据特定情况选择正确的统计检验方法。第13章介绍了一些中级和高级的统计方法。建议你最后再阅读这一章，因为它以上述章节中介绍过的概念为基础。

Liwen Vaughan(**邱黎雯**)

(京)新登字 130 号

内 容 简 介

本书深入浅出地讲述了信息专业人员常用基本统计学方法——描述统计学、t 检验、χ^2 检验、方差分析、相关、回归、基本非参数统计检验的理论及应用，也对中、高级统计方法——多重回归分析和线性关系模型(LISREL)作了介绍。本书的突出特点是，着眼于统计学方法应用的完整过程，着重于在对统计学基本逻辑的理解、统计学方法的选择和统计分析结果的解释上进行形象化描述，而不陷入技术细节和繁琐的数学公式之中，使读者感到轻松、实用。本书适合大专院校相关专业师生阅读和所有用到统计学方法的其他专业人员参考使用。

Third printing, July 2005

Statistical Methods for the Information Professional: A Practical, Painless Approach to Understanding, Using, and Interpreting Statistics

 Published by Information Today, Inc, 143 Old Marlton Pike, Medford, New Jersey 08055.

目　录