全国统计教材编审委员会“十一五”规划教材

统计学：从数据到结论

第二版

★ 吴喜之 编著

中国统计出版社
China Statistics Press

全国统计教材编审委员会“十一五”规划教材

统计学：从数据到结论

第二版

★吴喜之 编著

中国统计出版社
China Statistics Press

(京)新登字 041 号

图书在版编目(CIP)数据

统计学:从数据到结论/吴喜之编著. —2 版.
—北京:中国统计出版社,2006.9
ISBN 7—5037—4996—2

Ⅰ.统…
Ⅱ.吴…
Ⅲ.统计学
Ⅳ. C8

中国版本图书馆 CIP 数据核字(2006)第 104408 号

统计学:从数据到结论

作　　者/吴喜之
责任编辑/吕　军
装帧设计/艺编广告
出版发行/中国统计出版社
通信地址/北京市西城区月坛南街 57 号　邮政编码/100826
办公地址/北京市丰台区西三环南路甲 6 号
电　　话/邮购(010)63376907　书店(010)68783172
印　　刷/河北天普润印刷厂
经　　销/新华书店
开　　本/787×1092mm　1/18
字　　数/290 千字
印　　张/20.25
印　　数/1—5000 册
版　　别/2006 年 10 月第 2 版
版　　次/2006 年 10 月第 1 次印刷
书　　号/ISBN 7—5037—4996—2/C・2110
定　　价/38.00 元

出版说明

“十一五”时期是继续深化教育改革，加强素质教育，努力建设有利于创新型科技人才生长的教育培训体系的关键时期。为了更好地培育统计创新型科技人才，适应统计教育培训的新形势，全国统计教材编审委员会制定了《“十一五”全国统计教材建设规划》(以下简称规划)。规划坚持“以人为本”的科学发展观，坚持统计教育与实践相结合，坚持统计教育同国际接轨，坚持培养创新型的统计人才的指导思想，编写符合国民经济发展需要和统计事业发展需要的统计教材。

这批教材是在深入分析统计教育形势和统计教材建设发展状况，总结多年来统计教材建设经验的基础上，本着以建设本科统计教材为主的方针，积极探索研究生层次的统计教材，力争使规划统计教材的编写做到层次分明，有针对性和实用性。建设精品教材，是编委会自成立以来就孜孜以求的目标。考虑到统计教材建设的实际情况，“十一五”期间，本科教材主要以修订为主，对以往规划统计教材中使用面广，得到广大教师和学生普遍认可的教材组织了修订。修订后的教材，淘汰了过时的内容和例子，增加了计算机操作和大量的案例，编写手法也做了一定的调整，在实用性、可操作性等方面有了较大的改进。

近年来，我国现代化建设快速发展，高等教育规模持续扩大，尤其是研究生教育规模的扩大，使得高等学校研究生统计教学工作面临着许多新情况、新问题，任务艰巨。因此，必须坚持科学发展观，在规模持续发展的同时，把提高研究生统计教学质量放在突出的位置，培养全面发展的创新型的统计人才。教材是统计教学的载体，建设高质量

的研究生层次的统计教材是统计教育发展的需要。因此，编委会在“十一五”期间对研究生的统计基础课教材做了些有益的探索。根据《规划》的要求，这批教材主要采取招标和邀请的方式组织有关院校的专家、学者编写。

值得特别提出的是，在这批教材中，有《非参数统计》、《概率论与数理统计》、《经济计量学教程》、《医学统计》、《应用时间序列分析》、《多元统计分析》、《统计学》7种教材入选国家教育部组织编写的“普通高等教育‘十一五’国家级规划教材”，更加充实和完善了“十一五”期间统计教材的建设。

为了便于教学和学习，这批教材里面包含了与之相配套的《学习指导与习题》，使得这批教材在编辑出版上形成了比较完整的体系。我们相信，这批教材的出版和发行，对于推动我国统计教育改革，加快我国统计教材体系和教材内容更新、改造的步伐，打造精品教材，都将起到积极的促进作用。

限于水平和经验，这批教材的编审、出版工作还会有缺点和不足，诚恳欢迎教材的使用单位、广大教师和同学们提出批评和建议。

全国统计教材编审委员会

2006年6月

前 言

有什么在本书中等待着你们去发现，去探讨，去欣赏呢？当然不是数学公式和定理定义的堆砌，也不是和枯燥的公文报表相关的政府工作的培训。这是一门充满了哲学韵味的认识世界的学问。

不知读者们是否意识到，统计已经渗入到人们的社会、生活、工作等各个领域。每天新闻媒介报道的各个方面都离不开各种统计数据和各种分析与预测。人们可能对于这些统计内容觉得习以为常，也可能会有一些好奇或神秘感。由于国情不同，统计的地位与人们对统计的看法也不同。在发达国家，一般民众觉得统计学和数学类似，是一门高不可攀但极易找到满意工作的学问。在中国，又有一些人认为统计就是处理政府报表的职业。但自从中国向世界开放之后，越来越明确的一点是，没有什么学科或领域能够真正离开统计。

以应用为目标学习统计，究竟是为什么？是为了流利地背诵一大堆定义、概念和抽象的名词和术语吗？是为了学习如何进行推导和证明一些复杂的定理和公式吗？这些问题不仅学生会思考，更重要的是统计教师要思考。本书的目的是希望读者在学习之后，能够知道实际中哪些是统计问题，最好能够自己解决一部分统计问题，即使不能解决也知道能够在哪里查到答案和向谁请教。知识固然重要，更重要的是通过学习获得解决和处理问题的能力。

学习并不总是一个令人生畏或至少成为某种负担的过程。人们学会走路、说话、骑车、下棋、打球等大都是在一种乐趣中进行的。为什么涉及到日常生活的每一个方面的统计就不能和看侦探小说那么引人入胜呢？其实任

何一门科学，都有其趣味性；而只有把科学研究当成游戏的人才会真正成为大师。这门课并不想使读者都成为统计学家，而仅仅想让读者如同学会使用电脑、手机，学会辩论、上网或讨价还价那样愉快地认识或理解在人生中无法躲开的统计。

本书由浅入深地把统计最基本和最有用的部分在这么一本不厚的教科书中完整地介绍给读者；而且让读者可以边学习，边着手用统计软件处理数据。篇幅大、语言啰嗦的教材对读者是个负担，不但浪费了资源，也抓不住要领。因此，作者力图惜墨如金；既节省篇幅，又要把该解释的全部说清。希望读者慢慢咀嚼，不必图快。

很少有一本统计教材包括像本书那么多的统计内容。我觉得，这些内容本来并不深奥；只是其貌似复杂的数学工具把它搞成阳春白雪；再加上强调数学推导的教学方式，使得统计显得高不可攀。本教材要还这些统计应用以其本来面目。使得统计变成人人都能够基本上理解和掌握的有用工具。多数使用计算机的人都不是计算机专业的；多数开汽车的都不会修汽车；但这对他们毫无妨碍。难道不会推导或背诵与统计有关的数学公式就不能应用统计这个工具了吗？

本书每一章的主要部分是用日常语言来引进和解释一些概念，如果可能就通过例子来说明。如果不涉及应用，这部分就足够了。涉及应用的各章后面的小结中，有一部分是说明如何通过统计软件来处理本章的数值例子；这会给多数想要自己动手分析数据的读者以方便。小结的最后还展示了与概念及计算有关的一些数学公式，使那些精力充沛的读者能更深刻地理解内容。这种安排使得本教材能够适用于各种不同水平、不同要求的读者群体。本教材不仅可供没有学过概率论和数理统计的非统计专业的本科生和研究生使用，也可以供统计专业的本科生作为理解统计本来含义的教材使用（以代替不能满足需要的“描述统计学”等类课程）；它还可以为各领域的广大实际工作者作为应用各种统计方法的参考书。为读者可以使

用各种软件来进行分析，本书所涉及的所有电子版数据都有文本格式、SPSS 格式及部分 EXCEL 格式。

软件方面，本书主要使用 SPSS 和部分地应用 Excel。我们不可能介绍所有软件，也不可能介绍某个软件的所有细节。经验证明，只要把某个方法在某个软件的基本选项指出，学生就可以通过自己的经验（最多借助于帮助）来得到所需要的结果。在课本中罗列使用软件时的出现的各种对话（选项）框的做法对本课程完全没有必要。

在前计算机时代，几乎所有的统计教科书都给出了各种与分布有关的表格。但随着计算机的普及，所有统计软件（无论是商业的还是免费的）都给出了和各种分布有关的各种函数，把人们从繁琐而又不精确的查表中解放出来。目前很多国外的统计教科书都不再提供了既占用篇幅又比较粗糙的分布表。本书不准备提供任何和分布有关的表格。本书第四章会介绍如何使用软件来进行与概率分布有关的计算。

这个教材的全部内容曾作为非统计专业硕士和博士的课程分别在北京大学光华管理学院及中国人民大学讲授过，受到普遍欢迎。实践证明，这个课程前 15 章的内容完全能够轻轻松松地在一个学期（每周三个学时）中全部讲完。一些热心而又好奇的非统计背景的人士也曾读过本教材的全部内容，没有任何理解上的问题。当然，根据不同的教学对象和需要，有些章节可以完全不讲或少讲。

本书前面的章节，是对统计基本概念的介绍。而后面的部分则是更有针对性的一些统计模型和方法。一般传统统计学的课程包括前六章，或最多前九章的内容；而第十章到第十四章一般属于多元统计分析的课程内容；第十五章一般属于时间序列课程包含的内容；第十六章一般属于非参数统计课程的内容；第十七章介绍了生存分析；第十八章对指数进行了必要的介绍。目前大多数流行的统计应用都已包含在本教材内。

本书的编写是在国家统计局教育中心的建议和鼓励下产生，并得到其大力支持。本书还受到北京大学、中国

人民大学以及各兄弟院校老师和学生的鼓励和帮助。中国统计出版社一直关心着本书的写作和出版。SPSS 北京办事处的专家也一直积极对写作过程中出现的有关计算问题予以帮助。特别要指出的是敬爱的汪仁官老师又一次为我所写的统计教材进行了非常认真的审校，使我重新感受到做学生的幸福；中国统计界的老前辈茆诗松老师也热心地对本书提出了许多宝贵而又中肯的建议。他们的审校和建议使本书避免了许多错误和不妥之处。没有这些支持和帮助，本书是不可能面世的。谨在此对所有各方面表示衷心的感谢。

吴喜之

2003 年 6 月

第二版说明

本书的第一版发行不到一年，已经作为参考书或教科书在许多学校使用。各个学校的师生对本书提出许多宝贵的意见，并且指出了很多错误和不妥之处。他们的支持和鼓励，促使了本书第二版的诞生。

和第一版相比较，第二版对许多内容完全重新写过，还进行了一些调整，同时加强了对概念和方法的解释，使得该书更加容易理解。第二版还对例题和习题都做了很多修订和增减。此外，还增加了一些内容；除了基于SPSS的操作和输出结果的分析和解释之外，还增加了SAS软件的使用；特别是增加了关于如何通过免费的，功能强大的，需要自己动手写程序的R软件来理解概念及处理数据。R软件的代码公开及透明的优势是一些“黑匣子”式的傻瓜软件所无法比拟的。R软件是使用S语言来编程的(和S-plus的编程语言一样)；在其问世的不到10年的时间，已经成为国外统计研究生的首选软件。它有强大的网上支持系统。多数最新的统计计算方法，在进入商业软件之前，就已经以R语言的形式在R网站上免费提供了。建议使用本书的师生最好也使用R语言。在掌握R软件之后，对其他统计方向的学习和研究都会有很大的帮助，甚至会有一种到了自由天地的感觉。在R软件的帮助中有完整的入门材料和各种命令的意义和使用例子；因此本书没有自己编写关于R语言的附录来增加篇幅。

本书选择SPSS软件选项和SAS软件语句(或选项)的原则是容易理解和掌握；当然，由于编者知识有限，对于有些方法，没有找到(因此也无法提供)最合适、最方便或者最新的软件选项或模块；希望读者提出建议，使得再版

时予以弥补。

由于使用软件比查表更加方便和可靠，可能会有人说，你自己都不查表，为什么要教学生去查表呢？的确，编者除了在最初等的统计课教学过程中曾经涉及到少数统计表之外，从来都是使用软件。“己所不欲，勿施于人”，本书不提供任何统计分布表。希望有条件的读者尽量使用计算机，而不去查表。实际上，如果没有计算机的支持，很难对有一定规模的数据在任何统计方向进行较深入的分析。

有许多人（比如各层管理人员）不一定要进行第一线的实际数据计算，但为了理解手中关于本单位及有关方面信息的意义，为了更好地进行明白的决策，他们必须理解各种统计推断结果的意义。对这些人，不一定要求能够熟练使用软件，更不需要理解数学推导，但必须明白各种统计概念和方法以及输出结果的意义。相信本书能够对他们有所帮助。

作为教科书，本书内容对于每周两学时的课程似乎太多。但是，什么讲或者什么不讲可以根据学生的需要由老师自己安排。实际上，对于任何课程，最好是由任课教师来决定讲哪些内容以及如何讲。因为他们最了解他们所面对的学生。教科书编者的思维方式不见得和讲课老师的一致，而老师最好按照自己的理解来讲述。一本好的教科书，应该给教师以较大的余地和自由。

希望读者继续对本书予以宝贵的支持和批评指正。

吴喜之

2006 年 3 月

目录

目录

目录

目录

目录

目录