HZ BOOKS
华章科技

本书是Stata入门必备书籍，全面、系统地讲解了软件的基础知识、数据访问及管理等内容，并配有大量来源于实践的案例

本书结构系统，讲解清晰，通俗易懂，实用性强，是统计人员、Stata初中级用户、医学及生物研究人员、数据管理员以及其他数据分析人员的必备参考书籍

大数据技术丛书

Data Analysis and Statistics in Stata

Stata数据统计分析教程

廉启国◎编著

机械工业出版社
China Machine Press

Data Analysis and Statistics in Stata

Stata数据统计分析教程

廉启国◎编著

机械工业出版社
China Machine Press

图书在版编目（CIP）数据

Stata 数据统计分析教程 / 廉启国编著. —北京：机械工业出版社，2015.4

ISBN 978-7-111-50028-5

Ⅰ. S… Ⅱ. 廉… Ⅲ. 统计分析－应用软件－教材 Ⅳ. C819

中国版本图书馆 CIP 数据核字（2015）第 084296 号

Stata数据统计分析教程

出版发行：机械工业出版社（北京市西城区百万庄大街22号 邮政编码：100037）

责任编辑：李华君　　责任校对：董纪丽

印　　刷：北京瑞德印刷有限公司　　版　　次：2015年5月第1版第1次印刷

开　　本：185mm×260mm　1/16　　印　　张：24

书　　号：ISBN 978-7-111-50028-5　　定　　价：69.00元

凡购本书，如有缺页、倒页、脱页，由本社发行部调换

客服热线：(010) 88378991　88361066　　投稿热线：(010) 88379604

购书热线：(010) 68326294　88379649　68995259　　读者信箱：hzjsj@hzbook.com

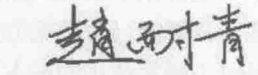
序

复旦大学公共卫生学院是国内最早开设卫生统计学课程的院校之一，也是最早提出用 Stata 软件进行教学的院校之一。该院从 1988 年起开设了统计软件课，通过教学实践，最终确定 Stata 软件是最为合适的教学软件。该软件功能强大，操作简便，输出结果针对性极强，并且其许多功能都是针对医学研究背景设计的。

从 2004 年开始，我们开始摸索新的教学方法，尝试以研究设计为主线，以数据类型为分类，以数据分析为教学目标，在统计学教学中融入统计软件 Stata 的基本操作，淡化统计计算公式的步骤，借助计算机辅助教学，强化统计学基本概念，强化学生数据分析的能力。

在学习了复旦大学公共卫生学院开设的卫生统计学课程后，本书作者廉启国熟悉并喜欢上了 Stata，且将 Stata 作为工作中数据分析的首选统计软件。Stata 软件的更新速度非常快，一般保持每两年一个大版本的更新速度。本书基于 Stata/MP 13.1，深入浅出地对 Stata 的常用功能及最新功能进行了介绍。本书的第 1 章介绍了 Stata 的入门知识，第 2 章至第 10 章介绍了 Stata 在数据访问、数据管理、数据呈现和数据分析 4 个重要方面的强大功能，第 11 章介绍了 Stata 的编程技巧。此外，本书还专门在第 12 章和第 13 章介绍了 Meta 分析和样本量估算的 Stata 实现。

我相信，此书的出版将为国内许多师生使用 Stata 提供很好的帮助。



复旦大学公共卫生学院

2015 年 1 月

前 言

为什么要写这本书

Stata是一款无与伦比的统计软件，它优雅、高效且易学。Stata诞生于1985年1月，第一个版本只有44个命令和175页文档。2015年是Stata诞生30周年，目前Stata已经发生了翻天覆地的变化，历经30个版本的迭代，已经升级至14.0版，支持矩阵编程和多核处理器，且提供20本合计超过11 000页的文档。Stata日趋完美，用户遍布200多个国家和地区，已成为最重要的统计软件之一。

作者是2004年在复旦大学读硕士研究生时，在赵耐青教授和张文彤博士的课程中接触并喜欢上Stata的，十年弹指一挥间，也积累了一些Stata的使用经验。一个偶然的机会，在机械工业出版社华章公司李华君编辑的建议下，开始尝试本书的撰写工作。写作的过程也是一个重新学习和梳理的过程，因为将自以为掌握的知识写出来并非易事。

Stata 8是Stata发展的一个重要里程碑，但是在生物统计学和医用统计学领域，有关Stata的书籍还不是很丰富，很多书籍也都是介绍Stata 8之前的软件，未能紧跟Stata的发展步伐。本书定位为Stata入门级的书籍，以Windows平台下Stata 13.1 MP版本为基准，围绕数据访问、数据管理、数据呈现和数据分析4个核心问题进行了详细阐述（并介绍了大量实用且免费的第三方命令）。此外，本书还介绍了Stata编程以便提高用户日常科研工作效率并介绍了工作流程以有助于用户更好地实施项目管理（含数据的加密和恢复）。最后，本书通过两章对近年来比较热门的Meta分析和日常科研中频繁使用的样本量估计进行了介绍。

读者对象

本书适合的阅读对象如下：

- Stata初中级用户
- 医务人员
- 医学/生物科研人员
- 数据管理员
- 生物统计师
- 其他行业有数据分析需求的人员

如何阅读本书

本书的每个章节都是一个相对独立的知识块，读者可以选择从头开始逐章阅读，如果有一定Stata基础和使用经验的话也可以根据目录跳转到感兴趣的章节。本书主要分为6部分，总计13章，另外还有5个附录。基本结构如下。

第一部分　软件入门（第1章）

第一部分介绍了Stata的基础知识，包括Stata软件概述（发展历史与版本选择、正确读写与文献引用、软件特点与优势）、Stata操作入门（安装与激活、启动与退出、操作方式、结果输出、自定义设置）、Stata使用基础（文件格式、变量类型、运算符、语法规则）、Stata升级扩展（软件升级、第三方命令的查找与安装、帮助文件及学习资源）。

第二部分　数据访问（第2章）

第二部分介绍了Stata在数据导入/导出方面的功能，包括Stata自带的数据导入和导出功能、数据导入的第三方命令（usespss）、数据转换软件（Stat/Transfer和EpiMate）、数据录入软件（EpiData Entry）。

第三部分　数据管理（第3～6章）

准确的数据是以后进行科学分析的基础，所以在进行数据分析之前，确保数据已被清理干净是非常必要的。数据管理是个非常枯燥的工作，但它是一切后续工作的基础。第三部分介绍了Stata在数据管理方面的强大功能（尤其是时间戳功能），包括数据清理、标签和注释、变量加工和数据集加工4章。

第四部分　数据呈现篇（第7～9章）

第四部分介绍了Stata在数据呈现方面的功能，包括统计描述、报表制作和图形绘制3章。Stata在数据呈现方面的功能非常强大，尤其是Stata的绘图效果堪称完美，建议对图形有特别要求的读者仔细阅读第9章的内容。

第五部分　数据分析（第10章）

Stata的数据分析非常完善，由于篇幅所限，本书只介绍了统计分析中最常用的一些Stata的功能，包括正态性检验和变换、t检验、方差分析、非参数检验、卡方检验、流行病学表格分析、相关分析、线性回归模型、logistic回归模型、等效性检验。在多元回归模型里特别增加了流行病学家和统计学家对自变量纳入的不同考虑。

第六部分　科研必备（第11～13章）

第六部分是为从事科研工作和数据处理工作的读者准备的，包括Stata编程基础、Meta分析、样本量与检验效能3章。建议对编程感兴趣的读者仔细阅读编程基础一章，医学/生物相关的专业人员需要特别关注的是Meta分析以及样本量与检验效能两章。

免责声明

本书中提及了大量的软件和服务，包括商业的（如：Stata、SAS、SPSS、Stat/Transfer、PASS、

Beyond Compare、FileWall、BitLocker、金山数据恢复软件、顶尖数据恢复软件）、免费的（如：EpiData、EpiMate、Power and sample size、uWall、魔方数据恢复、Eraser、金山快盘、有道云笔记、百度云盘、OneDrive、DropBox 和 Google Drive）和开源的（如：R）3 类。作者声明自己与这些软件和服务无任何利益关系，仅为提供更多信息。

此外，本书借鉴了大量的参考资料，并在所引用之处都尽可能地标注了文献出处，若有疏漏之处，作者在此深表歉意，并请及时联系我们，以便在后续版本中增加。

勘误和支持

由于作者的水平所限，经验不足，加之编写时间不尽宽裕，书中疏漏、错误之处在所难免，不妥之处恳请读者批评指正。

本书的勘误信息会发布在公卫人的 Stata 专版（网址：www.epiman.cn/forum-9-1.html），作者会在 Stata 专版中不定期更新书中的遗漏。当然，也欢迎读者将在阅读本书时遇到的疑惑、错误和建议在 Stata 专版中发帖提出。如果有些话题比较隐私，可以发送邮件至作者的邮箱（qglian@fudan.edu.cn），期待能够收到各位的真挚反馈。

致　谢

首先要感谢的是我的父亲和母亲以及岳父和岳母，你们在生活上的支持让我远离各种琐事，才可能将业余时间集中投入到本书的创作中。

其次要感谢我的妻子石慧颖女士，若不是她的持续鼓励和在背后默默的支持，按时交稿是一件无法完成的任务。特别感谢我刚满 3 岁的女儿廉孟洁，她每天都为家庭带来欢声笑语，也让我学会对世界始终保持一颗好奇的心。

还要感谢机械工业出版社华章公司李华君编辑的慧眼识才。因为平时的科研任务很重，所以虽然出书的想法产生已久，但若不是李编辑的鼓励，本书依然还在纸上谈兵的阶段。

本书内容参考了大量的专著、文章、文档和网站，此外，本书也介绍了众多 Stata 用户免费贡献的第三方命令，在此表示感谢。正因为有你们，Stata 社区才会如此繁荣。

最后，谨以此书献给 Stata 软件诞生 30 周年！

廉启国

2015 年 1 月于上海

目　录

第二部分 数据访问

第三部分 数据管理

第四部分 数据呈现

第五部分 数据分析