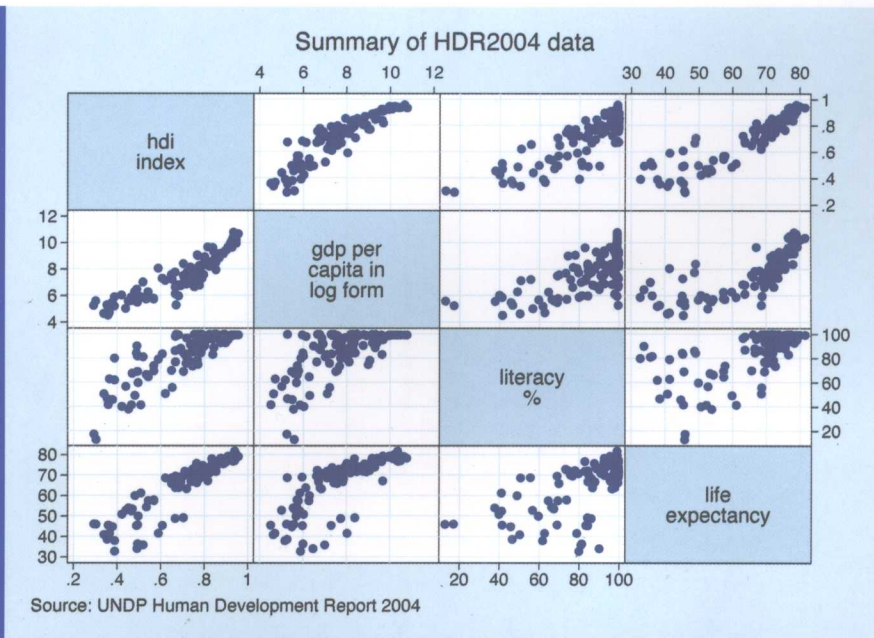


STATA

STATA 实用教程

王天夫 李博柏 著



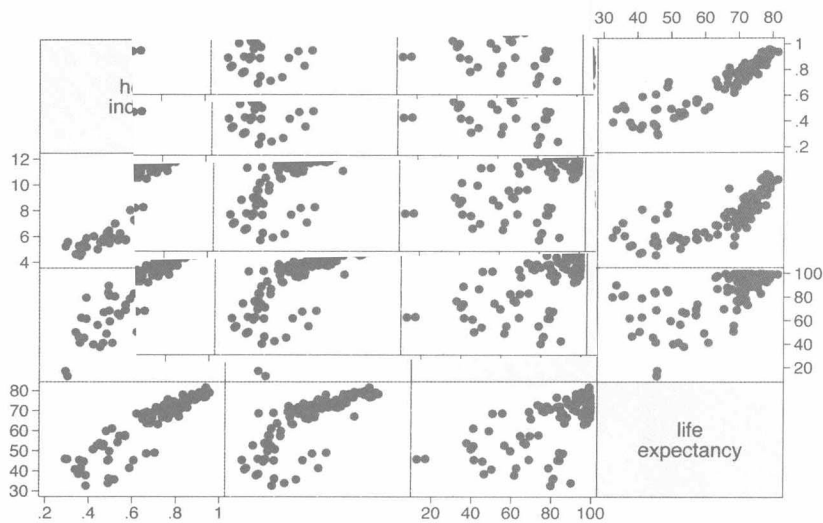
A GUIDE TO STATA

 中国人民大学出版社

STATA

STATA 实用教程

王天夫 李博柏 著



Source: UNDP Human Development Report 2004

A GUIDE TO STATA

中国人民大学出版社

· 北京 ·

图书在版编目 (CIP) 数据

STATA 实用教程/王天夫, 李博柏著.
北京: 中国人民大学出版社, 2008
ISBN 978-7-300-09639-1

- I. S...
- II. ①王... ②李...
- III. 统计分析-应用软件, STATA-教材
- IV. C812

中国版本图书馆 CIP 数据核字 (2008) 第 131508 号

STATA 实用教程

王天夫 李博柏 著

出版发行	中国人民大学出版社		
社 址	北京中关村大街 31 号	邮政编码	100080
电 话	010-62511242 (总编室)		010-62511398 (质管部)
	010-82501766 (邮购部)		010-62514148 (门市部)
	010-62515195 (发行公司)		010-62515275 (盗版举报)
网 址	http://www.crup.com.cn		
	http://www.ttrnet.com (人大教研网)		
经 销	新华书店		
印 刷	北京东君印刷有限公司		
规 格	185 mm×230 mm 16 开本	版 次	2008 年 9 月第 1 版
印 张	29.25 插页 1	印 次	2008 年 9 月第 1 次印刷
字 数	481 000	定 价	49.80 元

版权所有 侵权必究 印装差错 负责调换

前言

随着大规模数据与统计分析在实际学习和工作中的使用日益频繁，越来越多的人认识到能够熟练使用统计软件的重要性。很多时候，正是由于统计分析与计算手段的局限，导致了不少人对庞大复杂的数据资料望而却步。在我们看来，Stata 是一个功能强大而又相对简单易学的统计软件，值得向更多的人介绍和推广，让更多的人学会使用 Stata，提高人们对大规模数据的处理和分析的能力。

本书的目的

本书旨在将 Stata 软件介绍给更多的中文使用者。我们希望读者可以通过本书清楚地了解 Stata 是怎样将特定的数据处理与统计分析技术与计算机技术结合起来的，并在此基础上熟练使用 Stata 来解决学习、研究与工作中的实际问题与任务。

Stata 本身并不是写给特定用户群体的软件，因而，本书的读者群也没有特别的界定。任何在实际工作学习中需要使用数据处理与统计分析的学者、研究者、政府与公司工作人员或是学生均可以通过学习本书进入到 Stata 的领域当中。本书不区分读者的学科背景，也不区分读者的统计知识层次。同时，本书对于读者的计算机基础知识没有任何要求，能够使用 Windows 操作系统的读者就应该有足够的学习能力学习 Stata。我们希望不同需求的读者都能够通过本书学习 Stata；同时也希望本书可以起到抛砖引玉的作用，为有兴趣的读者提供一个继续学习和探索 Stata 的基础。

本书的内容

本书并不详细讲解各种统计知识，而更多地专注于介绍怎样通过使用 Stata 来实现数据处理与统计分析的任务。鉴于目前有关 Stata 软件的中文书籍匮乏，我们尽量在本书里涵盖了更多的内容。同时，结合多年学习使用 Stata 的经验心得，我们特别构思了整本书的结构，意在使读者在学习 Stata 的使用的过程中从我们的经验教训中受益。

本书的内容大致可以分为三个部分，各个部分分别对应不同的实际需要。第一部分包括第一章到第五章，介绍 Stata 的基础知识与基本技能，所有读者都需要对这部分的内容有所了解才能较好地运行 Stata。第二部分包括第六章与第七章，将读者引入实际工作与学习的任务中，介绍简单的统计分析方法。第三部分包括第八章到第十章，是扩展的部分，涉及更为复杂的任务，如绘图、数据管理、Stata 编程。

为配合以上的讲解，本书提供了大量的数据处理与统计分析例子。除个别模拟数据以外，书中使用的大部分数据是我们在学习与工作中常见的实际数据，而分析的程序也都是我们在自己的研究工作中积累起来的。这些数据和程序都将被上传到网上，供读者下载。读者可以通过对这些材料的学习与实际应用来提高自己对 Stata 的实际操作能力。

在介绍统计分析技术时，我们仅仅选择了初级的分析技术，旨在让读者熟悉 Stata 分析数据的使用方法。更多的统计分析的技术与统计模型的处理，我们将在另一本书中做详细的讨论。

Stata 软件更新很快。我们最早使用的版本是 Stata 4.0，后来随着 Stata 的更新而不断学习补充。现在，Stata 的最新版本已经是 10.0。本书的全部内容和例子都是基于 Stata 10.0 的。

本书的排版和体例

本书包含大量的 Stata 程序编码和统计结果，给排版工作带来了很大的困难。

为了保证排版和印刷质量，我们通过自己设计的 LaTeX (L^AT_EX) 编排格式来完成整本书的排版印刷。在这个过程中，我们设计采用了以下特别的编排格式，意在提高本书的可读性，方便读者学习和练习。

- 对于正文中出现的技术性英文字节，我们使用了不同的字体，使之与正文中的一般英文区别开来。Stata 命令、符号以及函数，使用的是 Courier New 打字机字体；而 Stata 数据中的变量名，使用的则是 *Times New Roman* 斜体字体。
- 对于所有实际运行的 Stata 命令及其结果，我们使用不同于正文的格式，同时利用文字印刷的深浅来区分 Stata 命令和 Stata 结果。
- 在很多地方，我们给出了命令或是命令选项的总结表格。同时在附录部分，我们给出了学习 Stata 不可不学的常用命令。
- 为了方便读者查询特定内容，我们专门制作了命令与符号索引，附在本书的最后。

本书的使用

与学习其他计算机软件一样，达到熟练使用 Stata 技术的唯一途径就是反复练习。读者可以使用本书提供的材料不断地练习，并在自己实际工作与学习的过程中有意识地使用 Stata。读者在使用 Stata 解决实际问题时，完全可以参照与模仿我们给出的 Stata 程序。本书所提供的 .do 命令程序文件都来源于解决实际统计问题的过程中，模仿这些范例，然后自己编写 .do 命令程序文件，是提高熟练程度的一个有效捷径。一旦熟悉 Stata 基本技能之后，读者可以挑战尝试使用更为复杂的 Stata 程序。

致谢

鼓励督促我们下定决心着手撰写此书的人是首先应当感谢的。他们是我们分别

在清华大学、北京大学以及其他单位和部门的朋友、同事与学生们。当然，最大的精神力量来自我们的家人。

每一本书的出版背后都有一个辛勤的编辑队伍。本书的适时出版更是凝结了编辑的汗水。潘宇女士与徐晓梅女士在本书的出版过程中付出了大量的精力与时间。

最后，祝读者学习 Stata 的过程是一段轻松愉快并收获丰富的经历！

王天夫 李博柏

2008年6月于北京成府路

目录

第一章 接触 Stata	1
1.1 Stata 的特点与优势	1
1.2 Stata 的计算环境	4
1.2.1 Stata 的版本类型	4
1.2.2 Stata 的目录结构	4
1.2.3 Stata 的工作界面	6
1.2.4 Stata 的命令输入方式	8
1.2.5 Stata 的文件类型	9
1.3 开始使用 Stata	11
1.3.1 设定界面	11
1.3.2 输入命令	13
1.3.3 调整工作目录	14
1.3.4 一些有用的 DOS 命令	15
1.3.5 内存设定	17
1.3.6 读入数据	19
1.3.7 浏览数据	19

1.3.8	退出 Stata	20
1.4	一个简单的 Stata 分析流程	21
1.4.1	了解数据结构	21
1.4.2	变量与变量取值	22
1.4.3	描述数据	24
1.4.4	列表	26
1.4.5	t-检验	27
1.4.6	相关系数	28
1.4.7	制图	28
1.4.8	模型拟合	30
第二章	使用 Stata 的基础	31
2.1	Stata 的命令结构	31
2.1.1	命令组成部分详解	31
2.1.2	命令及其缩写	33
2.1.3	变量与变量串	34
2.1.4	观察个案范围	35
2.1.5	命令举例	37
2.1.6	查找与重新输入已输入过的命令	37
2.2	使用 Stata 的运算结果	38
2.2.1	直接复制粘贴	38
2.2.2	结果记录文件的类型	39
2.2.3	生成与存储结果记录文件	39
2.2.4	打开与关闭文件	40

2.2.5	暂时关闭/打开记录文件	42
2.2.6	添加/覆盖记录文件	43
2.3	Stata 的系统参数	45
2.3.1	改变 Stata 内存的 Windows 变量	46
2.3.2	结果显示缓存规模	46
2.3.3	结果显示每行的长度	47
2.3.4	逐屏显示与分页符--more--	47
2.3.5	矩阵规模	48
2.3.6	改变结果记录文件的类型	48
2.3.7	Stata 更新	49
2.4	学习 Stata 的资源	50
2.4.1	Stata 出版的用法指南	50
2.4.2	Stata 的在线帮助	50
2.4.3	其他 Stata 参考书	51
2.4.4	Stata 用户群	51
2.4.5	Stata 用户程序安装	52
2.4.6	使用命令 net	52
2.4.7	网络资源	53
第三章	使用 Stata 数据文件	54
3.1	数据的结构	54
3.2	读取与生成数据文件	57
3.2.1	读取已有的 Stata 数据文件	57
3.2.2	读取文本格式的数据文件	60

3.3	在 Stata 中输入数据	66
3.3.1	使用数据编辑窗口	66
3.3.2	使用命令窗口	66
3.4	转换数据	67
3.5	数据的标签与注释	68
3.5.1	数据标签	68
3.5.2	数据的注释	69
3.6	数据文件的存储	71
3.6.1	数据文件的存储	72
3.6.2	数据文件的版本	73
3.7	数据的输出	73
3.7.1	输出成文本格式	73
3.7.2	输出成表格数据格式	74
3.8	数据文件使用命令的小结列表	74
第四章	数据处理与数据运算	76
4.1	变量与变量值	76
4.1.1	变量的名称	77
4.1.2	变量名的缩写	78
4.1.3	变量的取值类型	79
4.1.4	变量的测量类型	81
4.1.5	变量标签	82
4.1.6	变量数值标签	84
4.1.7	变量注释	87

4.2	变量取值的类型与存储格式	88
4.2.1	数字变量的取值与存储格式	88
4.2.2	字符变量及其存储格式	91
4.2.3	缺失数据	93
4.3	数据的显示	95
4.3.1	数字变量的显示格式	95
4.3.2	字符的显示格式	97
4.4	变量的运算	99
4.4.1	变量的更名与排序	99
4.4.2	数据的排序	101
4.4.3	Stata 中的表达式	103
4.4.4	Stata 中 if 引导的条件	106
4.5	生成新变量与变量赋值	109
4.5.1	使用 generate 命令	109
4.5.2	使用 replace 命令	112
4.5.3	使用 recode 命令	115
4.5.4	使用算术运算来重新赋值	117
4.6	连续变量转换成非连续变量	119
4.6.1	使用 autocode 函数	120
4.6.2	使用 group 函数	121
4.6.3	使用 recode 函数	121
4.7	虚拟变量的生成	122
4.7.1	使用命令 generate、replace 与 recode	123

4.7.2	针对类别变量使用 <code>tabulate</code> 命令	123
4.7.3	针对二分变量使用算术运算	125
4.7.4	使用快捷方法	126
4.8	数值变量与字符变量间的转换	128
4.8.1	从字符变量到数值变量	129
4.8.2	从数值变量到字符变量	131
4.9	系统变量	132
4.10	Stata 中的函数	134
4.10.1	Stata 中的常用函数	135
4.10.2	使用函数生成新变量	137
4.11	作为计算器的 Stata: 使用 <code>display</code> 命令	140
4.11.1	算术运算	140
4.11.2	函数运算	141
4.12	数据中的选择	143
4.12.1	保留或是删除观察个案	143
4.12.2	保留或是删除变量	144
第五章	编写命令程序文件	146
5.1	命令程序文件简介	147
5.2	命令程序文件的内容	150
5.2.1	版本注释	151
5.2.2	空行与编者注释	151
5.2.3	处理命令文件中的超长命令行	152
5.3	运行命令程序文件	154

5.3.1	结果显示与保存	155
5.3.2	错误及其排查	157
5.4	读取其他命令程序文件	159
5.5	命令程序文件举例 <code>profile.do</code>	160
第六章	数据描述、制表、比较与检验	162
6.1	查看数据	162
6.1.1	查看数据与变量的基本信息	162
6.1.2	列出数据与查看数据取值	164
6.1.3	用主干一枝叶图显示数据分布	167
6.2	描述性统计量	168
6.2.1	使用 <code>summarize</code> 计算描述性统计量	169
6.2.2	使用 <code>tabstat</code> 计算描述性统计量	171
6.2.3	使用 <code>ci</code> 计算置信区间	174
6.3	制表	175
6.3.1	单变量频次表	175
6.3.2	两变量列联表	177
6.3.3	多个表格与多维表格	180
6.3.4	显示描述性统计量的列联表	183
6.4	均值比较与 t -检验	185
6.4.1	单样本 t -检验	185
6.4.2	两个样本 t -检验	188
6.5	方差分析	190
6.5.1	一元方差分析	190

6.5.2	二元与多元方差分析	191
6.5.3	协方差分析	193
第七章	线性回归分析	195
7.1	相关分析	196
7.1.1	整行删除缺失数据的相关系数—— <code>correlate</code>	196
7.1.2	计算两两对应的相关系数—— <code>pwcorr</code>	197
7.2	回归分析	199
7.2.1	一元回归分析	199
7.2.2	阅读回归分析结果	199
7.2.3	多元回归分析	201
7.3	估算后的拟合值及其他统计量	204
7.3.1	模型拟合值与残差	204
7.3.2	其他用于诊断的统计量	206
7.3.3	估算方差与协方差矩阵以及膨胀因子	207
7.4	回归分析中的图形	208
7.4.1	一元回归分析中的散点图及拟合直线图	208
7.4.2	多元回归分析中的散点图矩阵与新增变量散点图	210
7.5	假设检验	213
7.6	变量转换	216
7.6.1	一个有关收入的例子	216
7.6.2	诊断正态分布的假设	218
7.6.3	收入的转换	220
7.6.4	重新带入回归分析	224

7.7	非连续性变量	225
7.7.1	两个组别的虚拟变量	226
7.7.2	多个组别的类别变量	230
7.8	交互作用	235
7.8.1	类别变量间的交互作用	235
7.8.2	类别变量与连续变量的交互作用	239
7.8.3	交互变量的自动生成——xi	241
7.9	逐步回归	242
7.9.1	向后逐步回归	243
7.9.2	向前逐步回归	244
7.10	小结与讨论	245
第八章	绘制图形	246
8.1	Stata 制图简介	247
8.1.1	主要的图形类型	249
8.1.2	图形的组成部分与制图命令的结构	249
8.1.3	图形处理	250
8.1.4	设定图形显示格式	251
8.1.5	寻求帮助	253
8.2	直方图	254
8.2.1	简单的例子	254
8.2.2	添加选项	255
8.2.3	添加更多的选项	257
8.2.4	多个直方图	259

8.3	扇形图	260
8.3.1	扇形图中的主要选项	261
8.3.2	多个变量的扇形图	262
8.3.3	根据其他变量分组的扇形图	264
8.3.4	观察数目的扇形图	267
8.3.5	三个提醒	268
8.4	条形图	269
8.4.1	条形图中的统计量与条形图的类型	269
8.4.2	单变量条形图	270
8.4.3	根据多个变量分组的单变量条形图	272
8.4.4	多变量条形图	274
8.4.5	横向条形图	277
8.4.6	多方向条形图——股票涨幅走势图	279
8.4.7	人口金字塔	282
8.5	散点图	285
8.5.1	改变散点图显示符号	286
8.5.2	处理重合的数据点	288
8.5.3	更多的显示选项	289
8.5.4	改变轴线刻度	291
8.5.5	多个散点图	293
8.5.6	矩阵散点图	294
8.5.7	散点连线图	295
8.6	直线图	297