

第2版

赢在职场

Stata

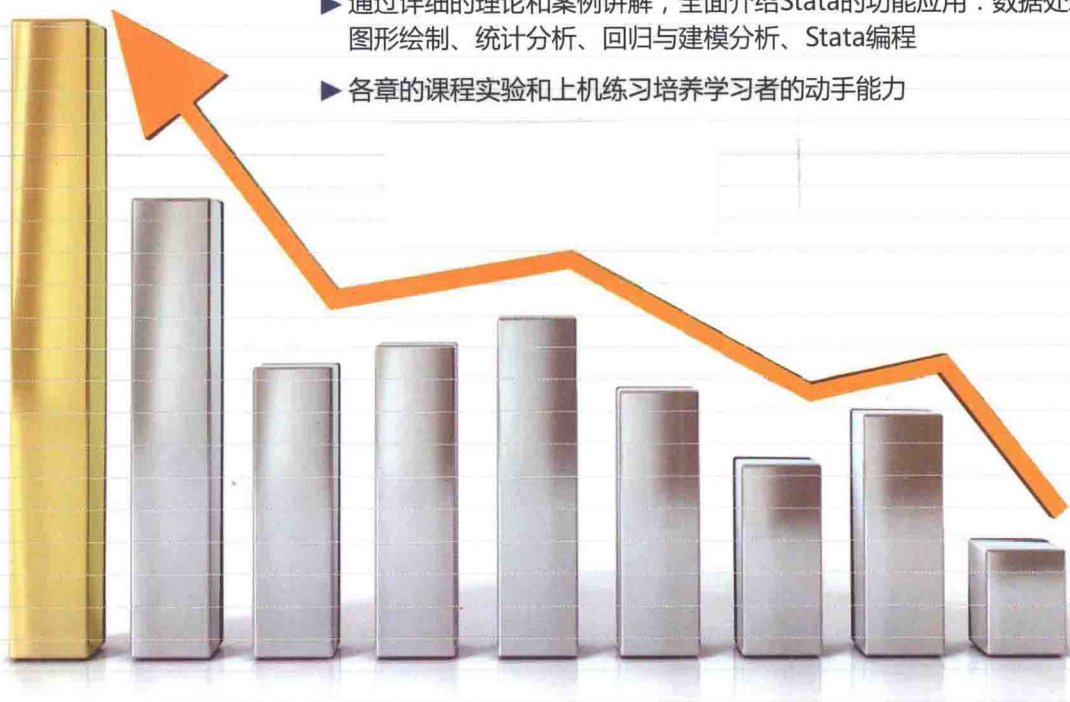
统计分析与应用



技术讲解+课程实验+上机操练

周广肃 等编著

- ▶ 通过详细的理论和案例讲解，全面介绍Stata的功能应用：数据处理、图形绘制、统计分析、回归与建模分析、Stata编程
- ▶ 各章的课程实验和上机练习培养学习者的动手能力



机械工业出版社
China Machine Press

第2版

Stata

统计分析与应用

周广肃 等编著



机械工业出版社
China Machine Press

图书在版编目 (CIP) 数据

Stata 统计分析与应用 / 周广肃等编著. —2版. —北京: 机械工业出版社, 2014.10

ISBN 978-7-111-48157-7

I. ①S… II. ①周… III. ①统计分析—应用软件 IV. ①C819

中国版本图书馆CIP数据核字 (2014) 第228446号

Stata 具有强大的数据处理和分析功能, 被广泛应用于统计学、经济学、生物学、医药学、社会学、人口学等领域, 功能十分强大, 现在普遍流行使用的版本为 Stata 10.0。

本书以通俗的语言讲解数据的基本处理、图形绘制、统计分析、回归与建模分析、编程等方面的内容, 讲解力求细致全面, 从而使读者熟悉和掌握 Stata 10.0 的各种功能操作。另外, 本书在每章的后面附有习题, 目的是培养读者的动手能力, 使读者在实际练习的过程中快速提高应用水平。本书作为实验教程, 十分注重内容的实用性, 不仅立足于典型案例进行教学安排, 还补充了所有案例背后对应的模型和原理, 方便读者巩固理论知识。

本书内容丰富, 结构清晰, 层次分明, 语言通俗易懂, 是一本较为实用的 Stata 实验教程。本书面向大中专院校经济管理专业及相关的社会科学专业学生, 特别是具有一定统计学和计量经济学基础知识的学生, 以及企事业单位和其他相关领域的科研工作人员。

Stata 统计分析与应用 (第 2 版)

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: 夏非彼 迟振春

印 刷: 中国电影出版社印刷厂

版 次: 2015 年 1 月第 2 版第 1 次印刷

开 本: 188mm × 260mm 1/16

印 张: 25.5

书 号: ISBN 978-7-111-48157-7

定 价: 59.00 元 (附光盘)

ISBN 978-7-89405-554-5 (光盘)

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) 82728184 88379604

购书热线: (010) 68326294 88379649 68995259

读者信箱: hzsj@hzbook.com

版权所有·侵权必究

封底无防伪标均为盗版

本书法律顾问: 北京大成律师事务所 韩光/邹晓东

前言

Stata 与 SAS、SPSS 统称为三大权威统计软件，被广泛应用于统计学、经济学、生物学、医药学、社会学、人口学等领域，功能十分强大。与其他软件相比，Stata 具有以下优势：操作简单，方便掌握；系统开放，及时更新；数据处理，功能强大；图形制作，精美优良等。Stata 软件具有 5 大功能，分别为数据的基本处理、图形绘制、统计分析、回归与建模分析、编程等。该软件是由 Stata 公司在 1985 年推出的，虽然现在最新的版本为 2009 年推出的 Stata 11.0，但是鉴于 11.0 版本还基本没有在我国推广和使用，所以本书所介绍的功能主要是通过 Stata 10.0 来实现的。

本书共 17 章，编写时先讲解各个实验案例的理论依据和模型，然后讲解该操作的通用命令模块，最后讲解操作的具体命令和命令执行结果。

- 第 1 章是 Stata 软件概述，主要介绍 Stata 软件的窗口、基本操作及主要功能模块。
- 第 2 章主要介绍数据的类型、压缩、转化、导入及整理。
- 第 3 章主要介绍 Stata 的图形制作，主要包括图形制作的基本命令与相关操作，直方图、散点图、曲线标绘图、条形图、饼图、箱线图的绘制方法，以及图形的保存、合并和修改。
- 第 4 章主要介绍单因素方差分析、多因素方差分析、协方差分析的相关实验内容。
- 第 5 章主要介绍单个总体的假设检验和两个总体的假设检验等相关内容。
- 第 6 章主要介绍小样本、大样本的 OLS 估计原理、操作方法和结果的解读，然后介绍约束回归和非线性最小二乘的相关实验内容。
- 第 7 章主要介绍虚拟变量的处理、经济结构变动的 Chow 检验、遗漏变量的检验、自变量数量的选择、极端数据的诊断与处理等。
- 第 8 章主要介绍多重共线性与逐步回归法、异方差检验与处理、内生性与 2SLS 等。
- 第 9 章主要介绍二值选择模型、多值选择模型、排序选择模型、条件和嵌套 Logit 模型等内容。
- 第 10 章主要介绍泊松回归模型、负二项和广义负二项回归模型、零膨胀回归模型等。
- 第 11 章主要介绍断尾回归模型、截取回归模型及样本选择模型。
- 第 12 章主要介绍时间序列的定义与扩展、相关图绘制与白噪声检验、移动平均滤波与指数平滑法、ARIMA 模型、SARIMA 模型、ARIMAX 模型、单位根检验、向量自回归模型、协整与向量误差修正模型、ARCH 族模型。
- 第 13 章主要介绍面板数据的基本操作、固定效应与随机效应模型等。
- 第 14 章主要介绍似不相关回归、多元回归模型、联立方程模型。
- 第 15 章主要介绍随机数的生成、重复抽样及自助法。
- 第 16 章是 Stata 编程基础，主要介绍基本概念与工具、程序文件的基本格式和程序



控制语句。

- 第17章是 Stata 综合案例分析，包括社会保障与经济增长关系实证研究、外部竞争环境不同的企业公司治理对绩效影响的实证研究、农民焚烧秸秆意愿的实证研究。

本书内容丰富，语言叙述清晰，结构简洁明了，通俗易懂，是一本较为实用的 Stata 实验教程。书中各章不仅通过详细的理论和案例讲解，而且还有与之配套的所有数据，在每章最后还附有一定数目的上机练习题，以供读者自行练习和应用，因此，读者只需按照书中介绍的步骤一步步地实际操作，就能完全掌握本书的内容。

本书面向大中专院校经济管理专业及相关的社会科学专业学生，特别是具有一定统计学和计量经济学基础知识的学生，以及企事业单位和其他相关领域的科研工作人员。由于本书编者尽量做到难易结合，所以本书既可以作为教材由老师负责讲解使用，也可以供读者自学使用。

本书主要由周广肃编写，参加本书编写工作的还有梁荣、田金秀、贾东永、凌佳、孙超逸、李龙、王华、李辉、刘峰、徐浩、李建国、马建军、唐爱华、苏小平、朱丽云、马淑娟、周毅、张浩、张乐、李大勇、许小荣、魏勇、王云等。本书在编写过程中吸收了大量的研究成果，在此对相关研究人员表示感谢。

编者力图使本书的知识性和实用性相得益彰，但由于水平有限，书中错误、纰漏之处在所难免，欢迎广大读者、同仁批评斧正。

编 者
2014 年 10 月

目 录

前言

第 1 章 Stata 软件概述	1
1.1 Stata 软件简介	1
1.2 Stata 窗口及基本操作	2
1.2.1 Stata 窗口说明	2
1.2.2 Stata 帮助系统	5
1.2.3 Stata 语法和命令	6
1.3 Stata 主要功能模块	8
1.3.1 数据处理	8
1.3.2 绘图	9
1.3.3 统计分析	9
1.3.4 回归与建模分析	9
1.3.5 编程	9
复习与习题	9
第 2 章 Stata 中的数据处理	11
2.1 数据的类型、压缩和转化	11
2.1.1 数据的打开	11
2.1.2 数据的类型与压缩	13
2.1.3 数据类型的转化	15
2.2 数据的导入	18
2.2.1 输入数据	19
2.2.2 使用已经保存的 Stata 数据	20
2.2.3 导入其他格式的数据	20
2.3 数据的整理	22
2.3.1 数据的标签与排序	22
2.3.2 数据的拆分	24
2.3.3 数据的合并	25
2.3.4 数据的长宽转换	27
复习与习题	28
第 3 章 Stata 中的图形制作	32
3.1 图形制作的基本命令与相关操作	32
3.1.1 图形制作的基本命令	32



3.1.2	图形制作的菜单选项	33
3.1.3	与图形绘制相关的基本操作	34
3.2	直方图、散点图和曲线标绘图的绘制	36
3.2.1	直方图的绘制	36
3.2.2	散点图的绘制	38
3.2.3	曲线标绘图的绘制	42
3.3	条形图、饼图和箱线图的绘制	44
3.3.1	条形图的绘制	44
3.3.2	饼图的绘制	47
3.3.3	箱线图的绘制	48
3.4	图形的保存、合并及修改	49
3.4.1	图形的保存和已存图形的打开	49
3.4.2	图形的合并	50
3.4.3	图形的修改	50
	复习与习题	51
第 4 章	Stata 与方差分析	55
实验 4.1	单因素方差分析	55
实验 4.2	多因素方差分析	58
实验 4.3	协方差分析	61
	复习与习题	64
第 5 章	Stata 与假设检验	66
实验 5.1	单个总体的假设检验	66
实验 5.2	两个总体的假设检验	71
	复习与习题	74
第 6 章	基本回归分析	76
实验 6.1	小样本的普通最小二乘分析	76
实验 6.2	大样本的普通最小二乘分析	84
实验 6.3	约束回归	87
实验 6.4	非线性最小二乘	91
	复习与习题	95
第 7 章	Stata 与模型的设定	97
实验 7.1	虚拟变量的处理	97
实验 7.2	经济结构变动的 Chow 检验	100
实验 7.3	遗漏变量的检验	102
实验 7.4	自变量数量的选择	107
实验 7.5	极端数据的诊断与处理	109
	复习与习题	111



第 8 章 Stata 与模型的修正	113
实验 8.1 多重共线性与逐步回归法	113
实验 8.2 异方差检验与处理	120
实验 8.3 内生性与 2SLS	127
复习与习题	132
第 9 章 Stata 与选择模型	134
实验 9.1 二值选择模型	134
实验 9.2 多值选择模型	144
实验 9.3 排序选择模型	149
实验 9.4 条件 Logit 模型	153
实验 9.5 嵌套 Logit 模型	157
复习与习题	164
第 10 章 计数模型	167
实验 10.1 泊松回归模型	167
实验 10.2 负二项和广义负二项回归模型	173
实验 10.3 零膨胀回归模型	179
复习与习题	184
第 11 章 样本选择模型与两步法	187
实验 11.1 断尾回归模型	187
实验 11.2 截取回归模型	192
实验 11.3 样本选择模型	197
复习与习题	203
第 12 章 基本时间序列分析	205
实验 12.1 时间序列的定义与扩展	205
实验 12.2 相关图绘制与白噪声检验	211
实验 12.3 移动平均滤波与指数平滑法	215
实验 12.4 ARIMA 模型	224
实验 12.5 SARIMA 模型	233
实验 12.6 ARIMAX 模型	238
实验 12.7 单位根检验	244
实验 12.8 向量自回归模型	248
实验 12.9 协整与向量误差修正模型	271
实验 12.10 ARCH 族模型	283
复习与习题	292
第 13 章 面板数据分析	297
实验 13.1 面板数据的定义、描述性统计与截面趋势图的绘制	297



实验 13.2 固定效应模型.....	301
实验 13.3 随机效应模型.....	303
实验 13.4 模型形式的选择——Hausman 检验	306
复习与习题.....	308
第 14 章 系统方程模型.....	309
实验 14.1 似不相关回归.....	309
实验 14.2 多元回归模型.....	314
实验 14.3 联立方程模型.....	316
复习与习题.....	322
第 15 章 自助法.....	324
实验 15.1 随机数的生成.....	324
实验 15.2 重复抽样.....	327
实验 15.3 自助法.....	334
复习与习题.....	340
第 16 章 Stata 编程基础.....	342
实验 16.1 基本概念与工具.....	342
实验 16.2 程序文件的基本格式.....	350
实验 16.3 程序控制语句.....	365
复习与习题.....	373
第 17 章 Stata 综合案例分析	376
17.1 综合案例一：社会保障与经济增长关系的实证研究	376
17.1.1 问题背景和数据描述	376
17.1.2 统计方法与 Stata 实现	377
17.1.3 结论	380
17.2 综合案例二：外部竞争环境不同的公司治理对绩效影响的实证研究	381
17.2.1 问题背景和数据描述	381
17.2.2 统计方法与 Stata 实现	383
17.2.3 结论	391
17.3 综合案例三：农民焚烧秸秆意愿的实证研究.....	392
17.3.1 问题背景	392
17.3.2 相关假设	393
17.3.3 数据来源及描述	394
17.3.4 统计分析与 Stata 的实现	394
17.3.5 结论	398

第 1 章 Stata 软件概述

Stata 软件作为现阶段较为流行的统计软件，其应用已经深入到经济科学、社会科学、行为科学、生物统计等等各个学科领域。本章作为全书第 1 章，主要介绍 Stata 入门的相关知识：首先将介绍 Stata 软件的基本情况，让用户对将要学习的软件有一个较为宏观的认识；接下来，将介绍 Stata 软件的窗口和基本操作命令，以熟悉基本操作；最后简要介绍下 Stata 的主要功能模块，这里我们也将对本书的内容进行一个概览，对于本书的章节安排和 Stata 的具体功能读者可藉此有一个直观的了解。

1.1 Stata 软件简介

Stata 软件具有强大的数据处理和分析功能，是现今较为流行的统计计量分析软件，它由 Stata 公司在 1985 年研制开发成功之后面市，到现在已经将近 30 年的历史了。在写作本书时其最新的版本为 2009 年刚推出的 Stata 11.0，但是鉴于 11.0 的版本还基本没有在中国推广和使用，所以本书所介绍的功能主要是通过 Stata 10.0 来实现的。

Stata 10.0 在安装时主要有 4 种版本，包括 Small（小型版）、IC（标准版）、SE（特殊版）和 MP（多处理器版）。用户可以在安装过程中自主进行选择，一般而言，SE 版已经能实现 Stata 的所有功能，MP 版与 SE 版相比，功能一致，但是运算速度更快。

Stata 与 SAS、SPSS 统称为三大权威统计软件，被广泛地应用于统计学、经济学、生物学、医药学、社会学、人口学等领域，功能十分强大。与其他软件相比，Stata 具有以下明显优势。

1. Stata 操作较为简单，方便掌握

Stata 为窗口编辑和程序操作的双操作方式软件，其操作的实现既可以通过点击菜单选项完成，也可以通过输入命令完成。Stata 的命令语句简洁明快，逻辑清晰，灵活方便，用户可以结合自身实际情况熟练应用。

2. Stata 是一个开放的软件系统

Stata 与其他统计软件相比，最大的优势就是可以不断地接收统计学和计量经济学的最新研究成果，保证与时俱进。众所周知，现在是一个知识爆炸的时代，计量和统计学发展迅速，新成果层出不穷。Stata 软件可以通过多种途径及时实现更新，保证用户可以将最新的理论成



果应用于自身的实践。这种开放性主要体现在三个方面：

- 用户可以到 Stata 官方网站 (<http://www.stata.com>) 上下载相关模块更新，或直接在命令窗口中输入“update”，在联网的情况下也可以实现更新。
- 使用 findit 命令，找到所需要的功能模块，并实现下载和安装。
- 下载由其他用户编写的 Stata 模块，通过菜单选项 Help -> SJ and User-written Programs 进行下载更新。当然，用户也可以自己编写程序，从而实现所需要的功能。

3. Stata 具有强大的数据分析功能

Stata 软件基本可以实现所有的统计与计量分析的功能。例如，在统计方面，可以实现数据的描述分析、方差分析、假设检验、主成分分析、聚类分析等；在计量分析方面，可以实现多种计量模型的应用，如单方程模型回归、联立方程模型回归、离散被解释变量模型、受限因变量模型、时间序列模型、面板数据模型、分位数回归模型等。Stata 软件这种统计与计量分析功能的有机结合，使得用户可以完成复杂性、综合性较强的各种分析和研究。

4. Stata 具有强大的图形制作功能

目前，图形分析仍然是各种分析必不可少的部分，具有强大的图形制作功能也是 Stata 被广泛接受的原因之一。利用 Stata 软件可以完成散点图、直方图、折线图、条形图、函数图等各种图形的制作，此外用户可以根据自身需要对图形进行修改，图形也可直接被图形处理软件和文字处理软件直接调用。这些功能使得 Stata 的画图功能方便有效，便于使用。

1.2 Stata 窗口及基本操作

本节将介绍 Stata 的窗口组成和基本操作，读者在这里将逐渐熟悉 Stata 的操作界面与方法。

1.2.1 Stata 窗口说明

Stata 软件安装完成后，运行 Stata，将会看到如图 1.1 所示的操作界面。

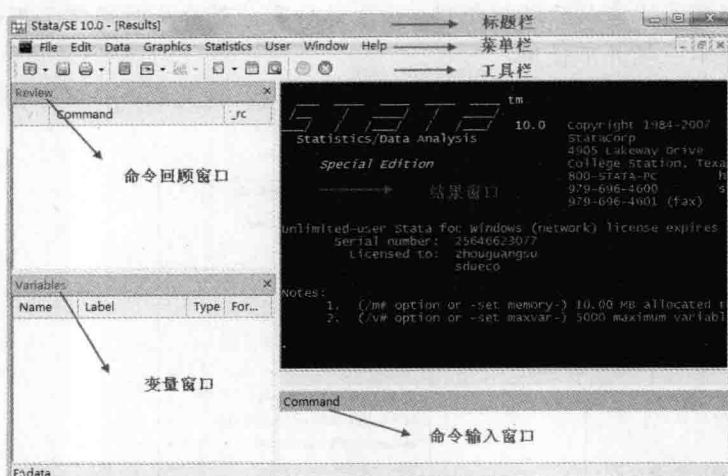


图 1.1 Stata 操作界面

从图 1.1 中可以看出, Stata 菜单栏主要包括 File、Edit、Data、Graphics、Statistics、User、Window、Help 这 8 个子菜单。由于 Stata 主要是通过命令进行操作, 所以这里只是简要介绍一下各个菜单的功能。

- File 的下拉菜单包括打开、保存、浏览、输入、输出以及打印等功能, 如图 1.2 所示。
- Edit 的下拉菜单包括数据的复制、粘贴等有关数据管理和设置的功能, 如图 1.3 所示。
- Data 的下拉菜单包括数据描述、数据编辑器、数据浏览、变量设置、矩阵运算等方面的内容, 如图 1.4 所示, 具体操作案例将在本书第 2 章中详细介绍。



图 1.2 File 的下拉菜单

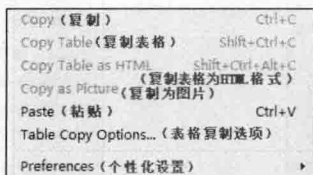


图 1.3 Edit 的下拉菜单



图 1.4 Data 的下拉菜单

- Graphics 是制图菜单, 制图种类主要包括散点图、点状图、柱状图、饼图等各种图形, 如图 1.5 所示, 具体内容将在本书的第 3 章中详细介绍。
- Statistics 是用来进行各种统计和计量分析的菜单, 如图 1.6 所示, 主要包括线性回归模型分析、时间序列分析、面板数据分析等方面的内容。这部分内容是本书的重



点, 将会在第 4 章以后详细介绍, 由于统计和计量分析时通常采用命令操作, 所以此部分菜单一般并不使用。

- User 主要是用来构建用户自己的菜单, 包括数据、图表和统计等方面的设置和操作, 如图 1.7 所示。



图 1.5 Graphics 的下拉菜单

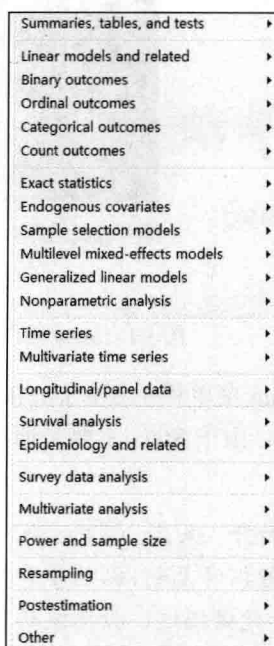


图 1.6 Statistics 的下拉菜单

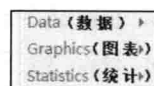


图 1.7 User 的下拉菜单

- Window 是对显示界面的操作, 包括对 Review、Results、Variables、Command 几大窗口的操作, 如图 1.8 所示。
- Help 为帮助菜单, 对用户未知的功能提供简单的帮助, 如图 1.9 所示。

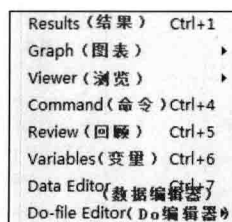


图 1.8 Window 的下拉菜单



图 1.9 Help 的下拉菜单

在 Stata 操作界面中, 最主要的部分由 4 大窗口组成, 分别是命令回顾窗口 (Review)、结果窗口 (Results)、变量窗口 (Variables) 和命令输入窗口 (Command)。

- 命令回顾窗口 (Review): 用于临时性存储已经执行过的命令语句。这些执行的命令语句主要包括两种: 一是直接从命令窗口中输入的命令, 另一种是通过窗口菜单



操作转化而成的命令。Review 窗口可以临时性存储自 Stata 本次运行到结束的所有命令，若 Stata 中途或最终被关闭，则所有的命令语句将会自动消失，若想永久保存这些命令，可以通过使用 log 命令或单击右键实现存储。在 Stata 运行过程中，可以重复使用显示在 Review 窗口中的命令，只需要左键单击命令，该命令将会重新显示在 Command 窗口中。

- 结果窗口 (Results)：用于显示命令执行结果的窗口。若命令执行的结果过长，则会在命令窗口的底端出现“more”这一字样，按下空格键或者回车键，可以浏览下面的内容；要想停止浏览，按“q”键或者工具栏中的图标则可。
- 变量窗口 (Variables)：用来显示变量名称和类型的窗口。同样点击某个变量名称，就可以将其显示在右方的命令窗口中。如果想要删除某个变量，在命令窗口中输入 drop 命令，鼠标左键单击相应的变量名称，然后按回车键即可完成操作。
- 命令输入窗口 (Command)：用户进行交互式程序操作的主要实现场所。在此可输入相关命令，然后按回车键即可。各种功能的相应命令将会在以后章节中详细介绍。

1.2.2 Stata 帮助系统

Stata 为用户提供了强大的帮助系统，新用户可以通过帮助系统的应用，更好地利用 Stata 完成自己所需要的功能和操作。Stata 的帮助系统主要由 Stata 手册、Stata 自带帮助和 Stata 网络帮助三部分组成，下面我们将一一进行介绍。

1. Stata 手册

Stata 手册是一本学习 Stata 使用的权威书籍，它按字母顺序排列出了 Stata 所有相关的命令。每一条命令不仅包括了其基本语法，还有具体操作实例、参考文献的说明，十分详细。但是对于中国用户来说，其缺点也不言而喻，Stata 手册数量庞大、价格昂贵，且用英语书写，对于英语欠佳的读者来说，使用极为不方便。

2. Stata 自带帮助

使用 Stata 自带帮助系统是最方便，也是最常用的方法。我们有两种方法可打开其自带的帮助系统：一种是从菜单栏中选择 Help 中的 Stata Command 选项，然后在弹出的对话框中输入所要查询的命令语句即可；另一种是更为常用的方法，即使用 help 命令，其语法格式为：

```
help [所要查询命令]
```

例如，我们想要知道非线性检验语句 testnl 的使用方法，就可以在命令窗口中输入 help testnl 这一命令，然后按回车键即可显示如图 1.10 所示的结果。



图 1.10 help 命令执行结果

从图 1.10 中可以看出：命令执行结果包括 Title——查询内容，Syntax——命令格式（语法），Description——命令描述，Options——选项，Examples——示例等主要内容，通过对查询结果的阅读，即可了解该命令的使用方法。

3. Stata 网络帮助

Stata 的网络帮助系统更为强大，用户可以在网上查找 Stata 还没有内置化的命令，实现自行安装。执行命令有两个，功能一致，具体如下：

```
findit [所查找命令], net
search [所查找命令], net
```

之后就可以按照提示命令进行下载安装了。

此外，用户还可以利用一些网站资源实现 Stata 的更好利用，以下是一些极好的资源：

- <http://www.stata.com>，这是 Stata 公司的官方网站。在此网站上，用户可以看到各种资源链接、新技术公告等各种各样有用的信息。
- <http://www.stata-press.com>，这是 Stata 出版社的单独网站，提供有关 Stata 出版物的信息。
- <http://www.stata-journal.com>，这是 Stata 电子杂志的官方网站。

1.2.3 Stata 语法和命令

掌握 Stata 的基本语法和命令，是熟练应用 Stata 做统计或计量分析的基础。首先，来了解一下 Stata 的基本命令语句的格式，具体形式如下：

```
[by varlist:] command [varlist] [=exp] [if exp] [in range] [weight]
[using filename] [, options]
```

“[]”中的内容表示可以省略的部分，因此我们可以看出，只有 command 是必不可少的，其他部分的内容可以根据自己的需要进行选择。下面详细介绍一下主要部分所代表的含义。



1. varlist

`varlist` 代表一个或多个变量，如果想表示多个变量，各个变量中间必须用空格隔开。对于已经存在的变量在调用时，用户可以使用一定的规则起到简化的作用，如使用“?”代表单个字符，使用“*”代表任意字符，使用“-”代表两个变量间的所有变量。然而对于新生成的变量来说，变量的名称不能进行简化。变量名称可以由字母、数字、下划线的组合构成，但是长度不能超过 32 位，当然名称不能以数字开头，也不能以单独的数字组合命名。

2. by varlist

`by varlist` 是分类操作的命令，表示对分类变量中不同的类别分别进行操作。下面将通过一个简单的例子演示一下这个命令的使用方法。

例题：利用美国 1978 年的汽车数据，计算一下美国汽车价格的总体平均值，以及进口车价格的平均值和国产车价格的平均值（数据文件：usaauto）。

计算总体价格的平均值较为简单，只需输入如下命令：

```
use auto, clear
summarize price
```

执行之后我们就可以得到如图 1.11 所示的结果。

```
. summarize price
```

Variable	Obs	Mean	Std. Dev.	Min	Max
price	74	6165.257	2949.496	3291	15906

图 1.11 总体价格均值

从结果显示中可以看出，这个命令求出了所有 74 个样本的价格均值为 6165.257（单位：美元）。

现在来分别计算进口车与国产车的价格均值，此时就可以用上刚才讲到的分类操作命令了，具体的命令语句如下：

```
by foreign: summarize price
```

这个语句的意思就是按照 `foreign` 这一分类变量将所有汽车划分为国产车与进口车，然后分别求出其价格的均值，结果如图 1.12 所示。

```
. by foreign: summarize price
```

-> foreign = Domestic					
variable	Obs	Mean	Std. Dev.	Min	Max
price	52	6072.423	3097.104	3291	15906
-> foreign = Foreign					
variable	Obs	Mean	Std. Dev.	Min	Max
price	22	6384.682	2621.915	3748	12990

图 1.12 分类价格均值



从结果中我们可以看出，国产车样本有 52 个，价格均值为 6072.423，进口车样本有 22 个，价格均值为 6384.682。

3. command

command 命令应用较多，也是此书介绍的主要内容，我们将在后面章节的相关内容中一一介绍。需要注意的一点是，在输入命令时可以在不产生歧义的前提下，尽量简写前几个字母。例如上面例题中的 summarize 命令，可以简写为 sum。

4. “=exp” 和 “if exp”

“=exp” 是赋值语句，“if exp” 是条件表达式。所涉及的函数有多种，在相关章节中将详细介绍。

5. in range

in range 主要用来表示观测值的区间，筛选出相应的样本，然后再执行相应的命令。表 1.1 列出了 in range 命令的使用方法。

表 1.1 in range 使用方法

表达式	含义	例子
in #	第#个观测值	in 6（第 6 个观测值）
in #1/#2	从第# ₁ 个观测值到第# ₂ 个观测值	in 2/6（第 2~6 个观测值）
in f/#	从第一个观测值到第#个观测值	in f/6（第 1~6 个观测值）
in #/l	从第#个观测值到最后一个观测值	in 6/l（第 6 至最后一个观测值）

6. weight

weight 表示的是观测值的权重，在加权最小二乘法中应用较多。权数的类型主要有抽样权数、重要性权数、频率权数、解析权数等。

7. options

options 选项在很多命令中都有，但是根据不同的命令其 options 选项也不尽相同，所以这一部分的具体内容将会在相关命令中介绍。

1.3 Stata 主要功能模块

Stata 软件主要有数据处理、绘图、统计分析、回归与建模分析、编程这 5 大功能，其相互配合，可以完成系统完整的数据分析和处理任务。

1.3.1 数据处理

在得到第一手数据之后要做的就应是对数据进行基本的处理，数据处理功能主要包括数