

万卷方法

USING STATA FOR PRINCIPLES
OF ECONOMETRICS

应用STATA学习

计量经济学原理 第4版

李·C·阿德金斯 (LEE C. ADKINS)

R·卡特·希尔 (R. CATER HILL)

著

曹书军 林健怡 译

曹书军 审校



重庆大学出版社
<http://www.cqup.com.cn>



WILEY

万卷方法

WILEY

USING STATA FOR PRINCIPLES OF ECONOMETRICS

应用STATA学习

计量经济学原理 第4版

李·C.阿德金斯 (LEE C. ADKINS) 著
R.卡特·希尔 (R.CATER HILL)

曹书军 林健怡 译

曹书军 审校

重庆大学出版社

USING STATA: FOR PRINCIPLES OF ECONOMETRICS, Fourth Edition . by LEE C. ADKINS and R. CARTER HILL

Copyright © 2011 by John Wiley & Sons, Inc. All rights reserved.

No part of this book may be reproduced or utilized in any form or by any means, electronic or mechanical, including photocopying, recording, or by any information storage and retrieval system, without permission in writing from the publisher. CHINESE SIMPLIFIED language edition published by CHONGQING UNIVERSITY PRESS, Copyright © 2015 by Chongqing University Press.

图书在版编目(CIP)数据

应用 Stata 学习计量经济学原理/(美)阿德金斯
(Adkins, L. C.), (美)希尔(Hill, R. C.)著;曹书军,
林健怡译. —重庆:重庆大学出版社, 2015. 9

(万卷方法)

书名原文: Using Stata for Principles of Econometrics

ISBN 978-7-5624-9441-6

I. ①应… II. ①阿…②希…③曹…④林… III.
①计量经济学—应用软件 IV. ①F224. 0-39

中国版本图书馆 CIP 数据核字(2015)第 200678 号

应用 Stata 学习计量经济学原理

(第 4 版)

(美)阿德金斯 希尔 著

曹书军 林健怡 译

曹书军 审校

策划编辑:雷少波 林佳木

责任编辑:文鹏 罗杉 版式设计:林佳木

责任校对:谢芳 责任印制:赵晟

*

重庆大学出版社出版发行

出版人:易树平

社址:重庆市沙坪坝区大学城西路 21 号

邮编:401331

电话:(023) 88617190 88617185(中小学)

传真:(023) 88617186 88617166

网址: <http://www.cqup.com.cn>

邮箱: fxk@cqup.com.cn (营销中心)

全国新华书店经销

重庆升光电力印务有限公司印刷

*

开本:787×1092 1/16 印张:32 字数:720 千
2015 年 10 月第 1 版 2015 年 10 月第 1 次印刷

印数:1—4 000

ISBN 978-7-5624-9441-6 定价:69.00 元

本书如有印刷、装订等质量问题,本社负责调换
版权所有,请勿擅自翻印和用本书
制作各类出版物及配套用书,违者必究

目 录

第1章 Stata 简介	1
1.1 启动 Stata	2
1.2 开启界面	2
1.3 退出 Stata	2
1.4 《计量经济学原理》的 Stata 数据文件	3
1.5 打开 Stata 数据文件	4
1.6 变量窗	5
1.7 数据描述与概要统计量	8
1.8 Stata 帮助系统	9
1.9 Stata 命令语法	11
1.10 文档存储	14
1.11 使用数据浏览窗	17
1.12 使用 Stata 制图	18
1.13 使用 Stata Do 文件	20
1.14 创建和管理变量	22
1.15 使用 Stata 密度函数	26
1.16 使用和展示标量	27
1.17 标量对话框	28
1.18 使用因子变量	29
第2章 简单线性回归模型	37
2.1 食品支出数据	38
2.2 计算概要统计量	40
2.3 创建散点图	41
2.4 回归分析	43
2.5 运用 Stata 获得预测值	50
2.6 估计非线性关系	51
2.7 用指示变量进行回归	59
第3章 区间估计和假设检验	76
3.1 区间估计	77
3.2 假设检验	78
3.3 p 值	81
第4章 预测、拟合优度和建模	93
4.1 最小二乘预测	94
4.2 拟合优度的度量	96
4.3 缩放和转换数据的影响	97

4.4	残差分析	101
4.5	多项式模型	103
4.6	估计对数线性工资方程	108
4.7	双对数模型	113
第5章	多元回归模型	122
5.1	模型举例	123
5.2	最小二乘预测	125
5.3	抽样的精确度	125
5.4	置信区间	126
5.5	假设检验	128
5.6	多项式方程	130
5.7	交互作用	133
5.8	拟合优度	135
第6章	多元回归模型:更多推断	141
6.1	F 检验	142
6.2	非样本信息	149
6.3	模型设定	151
6.4	数据质量差,共线性与不显著	159
第7章	使用指示变量	168
7.1	指示变量	169
7.2	应用指示变量	173
7.3	线性概率模型	181
7.4	处理效应	183
7.5	双重差分估计	188
第8章	异方差	199
8.1	异方差的性质	200
8.2	检测异方差	201
8.3	异方差一致标准误差	206
8.4	广义最小二乘估计量	208
8.5	线性概率模型的异方差	212
第9章	时间序列数据回归:平稳变量	219
9.1	引言	220
9.2	有限分布滞后	225
9.3	序列相关	226
9.4	序列相关的其他检验	230
9.5	序列相关误差估计	232
9.6	自回归分布滞后模型	236
9.7	预测	243
9.8	乘数分析	248

9.9 附录	250
第 10 章 随机解释变量和矩估计	261
10.1 工资方程的最小二乘估计	262
10.2 二阶最小二乘	263
10.3 用多余工具变量进行 IV 估计	267
10.4 内生性的豪斯曼检验	271
10.5 测试多余工具的有效性	274
10.6 弱工具变量的检验	275
10.7 CRAGG-DONALD F 统计量的计算	278
10.8 模拟实验	279
第 11 章 联立方程模型	295
11.1 松露的供给和需求	296
11.2 估计简约形式方程	296
11.3 松露需求的 2SLS 估计	297
11.4 松露供给的 2SLS 估计	301
11.5 鱼的供给与需求	302
11.6 鱼的价格和数量方程的简约形式	303
11.7 鱼需求的 2SLS 估计	304
11.8 2SLS 替代形式	305
11.9 蒙特卡洛模拟结果	309
第 12 章 时间序列数据回归:非平稳数据	320
12.1 平稳数据与非平稳数据	321
12.2 伪回归	326
12.3 平稳性的单位根检验	328
12.4 整合与协整	334
第 13 章 向量误差修正和向量自回归模型	341
13.1 向量误差修正(VEC)和向量自回归(VAR)模型	342
13.2 估计向量误差修正(VEC)模型	342
13.3 估计向量自回归(VAR)模型	346
13.4 脉冲响应与方差分解	351
第 14 章 时变波动率与自回归条件异方差(ARCH)模型	357
14.1 自回归条件异方差(ARCH)模型和时变波动率	358
14.2 估计、检验和预测	359
14.3 扩展	363
第 15 章 面板数据模型	370
15.1 微观计量面板数据	371
15.2 混合数据模型	372
15.3 固定效应模型	373

15.4	随机效应估计	382
15.5	回归方程组	389
15.6	混合模型	397
第 16 章	定性和受限因变量模型	408
16.1	二元因变量模型	409
16.2	二值选择的 Logit 模型	417
16.3	多项 Logit	423
16.4	条件 Logit	427
16.5	有序选择模型	431
16.6	计数模型	434
16.7	删失数据模型	435
16.8	选择偏差	441
附录 A	基本数学工具	454
A.1	Stata 的数学及逻辑运算符	455
A.2	数学函数	455
A.3	Generate 命令的扩展	455
A.4	计算器	456
A.5	科学计数法	456
A.6	数值的求导和积分	457
附录 B	概率论基础知识	460
B.1	Stata 概率函数	461
B.2	二项分布	462
B.3	正态分布	462
B.4	t 分布	464
B.5	F 分布	466
B.6	卡方分布	467
B.7	随机数	469
附录 C	统计推论回顾	476
C.1	检查 HIP 数据	477
C.2	使用模拟数据	479
C.3	中心极限定理	482
C.4	区间估计	484
C.5	检验正态总体均值	487
C.6	检验正态总体方差	489
C.7	检验两个正态总体均值相等	490
C.8	检验两个正态总体方差相等	492
C.9	正态性检验	493
C.10	极大似然估计	494
C.11	核密度估计	494

第 1 章

Stata 简介

本章概要

- 1.1 启动 Stata
 - 1.2 开启界面
 - 1.3 退出 Stata
 - 1.4 《计量经济学原理》的 Stata 数据文件
 - 1.4.1 工作目录
 - 1.5 打开 Stata 数据文件
 - 1.5.1 use 命令
 - 1.5.2 利用工具栏
 - 1.5.3 调用互联网文件
 - 1.5.4 本书资料的互联网地址
 - 1.6 变量窗
 - 1.6.1 使用数据编辑器修改变量标签
 - 1.6.2 使用数据工具修改变量标签
 - 1.6.3 使用变量管理器
 - 1.7 数据描述与概要统计量
 - 1.8 Stata 帮助系统
 - 1.8.1 使用关键词搜索
 - 1.8.2 使用命令搜索
 - 1.8.3 打开一个对话框
 - 1.8.4 Stata 操作手册中的完整文档
 - 1.9 Stata 命令语法
 - 1.9.1 summarize 的语法例解
 - 1.9.2 使用检阅窗口 (Review window) 学习命令语法
 - 1.10 文档存储
 - 1.10.1 复制与粘贴
 - 1.10.2 使用日志文件
 - 1.11 使用数据浏览窗
 - 1.12 使用 Stata 制图
 - 1.12.1 直方图 (Histograms)
 - 1.12.2 散点图 (Scatter Diagrams)
 - 1.13 使用 Stata Do 文件
 - 1.14 创建和管理变量
 - 1.14.1 创建 (生成) 新变量
 - 1.14.2 使用表达式构建 (生成) 器
 - 1.14.3 删除和保留变量与观测值
 - 1.14.4 使用算术运算符
 - 1.14.5 使用 Stata 数学函数
 - 1.15 使用 Stata 密度函数
 - 1.15.1 累积分布函数
 - 1.15.2 逆累积分布函数
 - 1.16 使用和展示标量
 - 1.16.1 标准正态 *cdf* 示例
 - 1.16.2 *t* 分布示例
 - 1.16.3 计算标准正态分布的百分位数示例
 - 1.16.4 计算 *t* 分布的百分位数示例
 - 1.17 标量对话框
 - 1.18 使用因子变量
 - 1.18.1 使用逻辑运算符创建指示变量
 - 1.18.2 使用 tabulate 创建指示变量
- 关键术语
第 1 章 Do 文件

1.1 启动 Stata

Stata 有多种启动方式。第一种方法,可以双击如下桌面图标:

这是 Stata/SE Release 11 的桌面快捷图标,早先的版本除了数字不同外,其余外观类似。当然,你也可以从 Windows 菜单点击进入,路径为开始 > 所有程序 > **Stata 11**。



第二种方法就是直接在 Stata 文件夹里找到扩展名为 ***.dta** 的文件,双击之。

1.2 开启界面

Stata 打开之后通常会出现如下不同标题的窗口:

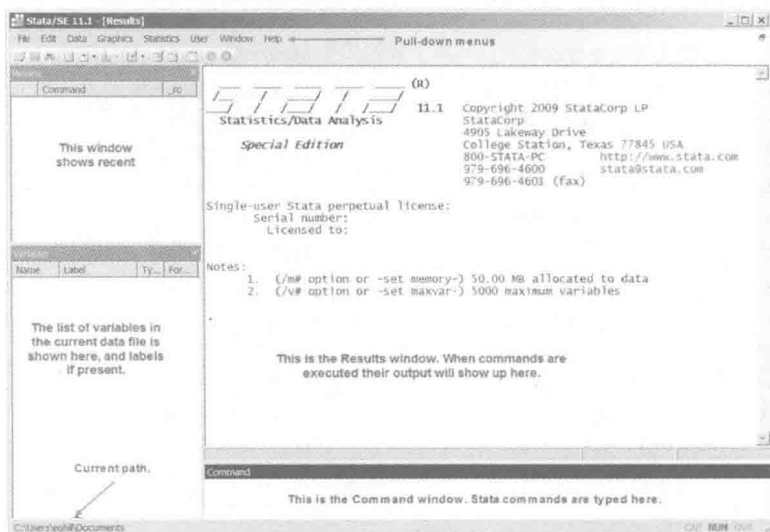
Command—命令窗,在该窗口输入 Stata 命令

Results—结果窗,本窗口展现命令的结果以及报错信息

Review—检阅窗,本窗口记录最近执行的命令清单

Variables—变量窗,本窗口列示数据变量名以及标签名称(如果有)

顶上一行是 Stata 的下拉菜单。其下一行从左起是 Stata 的工作路径,依次可以完成文件的打开、保存、打印等一系列操作,便于我们迅速地切换。

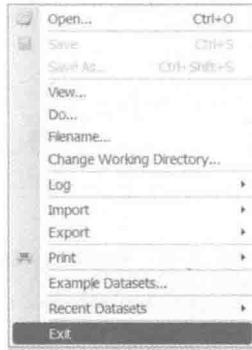


1.3 退出 Stata

退出 Stata 时可以点击 **File** 按钮。



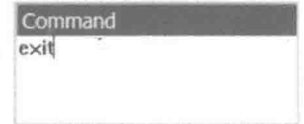
然后在下拉菜单里选择 **Exit** 按钮。



我们将使用 **File > Exit** 的方式表示这样的序贯点击命令。另外一种方式则是简单地在 **Command** 窗口输入：

```
exit
```

并回车。



1.4 《计量经济学原理》的 Stata 数据文件

Stata 数据文件的扩展名为 ***.dta**，此类文件只有 Stata 程序才能打开。如果锁定 ***.dta** 文件，并双击之，一样能开启 Stata。

本书中作为配套《计量经济学原理(第四版)》(*Principles of Econometrics, 4th Edition*，后文简称 *POE4*) 的 Stata 应用教材，原书里的案例数据都被转成 Stata 格式。这些数据适用于 Stata 9 及其后的版本，并且可以在下列网址中下载。

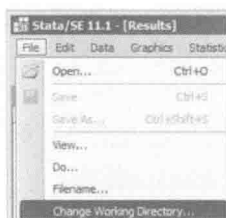
1. John Wiley & Sons 的网站：<http://www.wiley.com/college/hill>。你可以将那本书的所有数据集下载到自已电脑中。

2. 本书的数据以及其他资源可以到作者的网站上下载，网址：<http://www.principlesofeconometrics.com>。

3. 个别数据以及其他资料可以到 Stata 公司的网站上下载，网址：<http://www.stata.com/texts/s4poe4/>。

1.4.1 工作目录

一般可以把数据放入较为便利的目录。在本书中，我们将使用子目录 **c:\data\poe4stata** 文件夹来存放数据与结果。在实验室里，你的存储设备如果是闪存或移动硬盘，它们的容量要能够存储这些数据和文件。你一样可以在这些设备里以 **X:** 作为设备路径来设定子目录 **X:\DATA** 或者 **X:\POE4** 等，以便于处理。



还可以通过下拉菜单 **File > Change Working Directory** 来改变工作目录,然后在弹出的对话框里选择你的目标路径,并点击 OK。

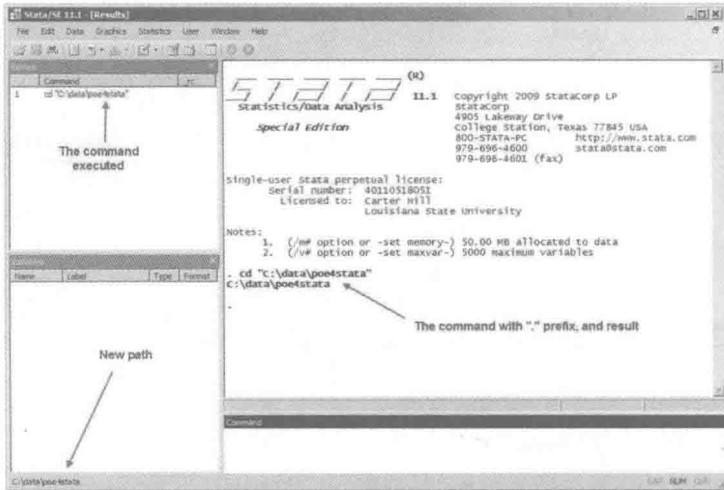
Stata 还可以在 **Command** 窗口输入:

```
cd "C:\data\poe4stata"
```

并回车。

结果如下所示:

```
Command
cd "C:\data\poe4stata"
```



注意一下,这个命令在 **Results** 窗口重现一遍,同时也出现在 **Review** 窗口。新的目录路径也出现在屏幕的左下方。

1.5 打开 Stata 数据文件

有多种方式可以打开或者导入 Stata 数据文件。

1.5.1 use 命令

打开 Stata,首先将工作目录切换到你的数据文件存放处。在 **Command** 窗口输入 `use cps4_small`,并回车。

```
Command
use cps4_small
```

如果你已经打开一个数据文件,并打算做某些改动,Stata 会做如下报错:

```
. use cps4_small
no; data in memory would be lost
r(4);
```

此时如果点击图框中的 `r(4)`,系统会弹出一个观察窗并在其中给出报错信息,大部分情况下这些信息是有帮助的。关闭该窗口只需点击其上的 X。

这里的告警信息能够有效阻止你丢失需要保留的数据。这种情况下,你可以保留早先的数据文件,或者直接输入命令:

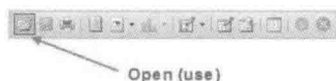
```
clear
```

需要注意的是,clear 命令将会清空 Stata 的系统内存数据信息。你也可以在打开数据的同时清空内容,方法是输入:

```
use cps4_small, clear
```

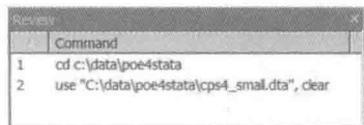
1.5.2 利用工具栏

我们还可以在 Stata 工具栏上点击 **Open (use)** 图标按钮打开 Stata 数据文件。



找到你要打开的文件,选之并点击 **Open**。在 **Review** 窗口,同时会出现隐含的 Stata 命令:

```
use "C:\data\poe4stata\cps4_small.dta", clear
```



在 Stata 里使用 `use` 命令可以打开数据文件,具体步骤见下图。数据文件的开启路径如引号内所示。具体路径必须加引号。“**clear**”选择项意味着此前内存中的所有数据同时被清空。

1.5.3 调用互联网文件

Stata 提供了足够多的互联网资源。本书的 Stata 数据文件存储地址为 <http://www.stata.com/texts/s4poe4/>。下载数据时,比如下载 `cps4_small.dta`,先要保存先前的数据,然后清空内存,然后在 **Command** 窗口输入如下命令:

```
use http://www.stata.com/texts/s4poe4/cps4_small, clear
```

数据一旦被下载到本地,就可以点击 **File > Save as**,然后在弹出的结果对话框里填入需要保存的文件名。

1.5.4 本书资料的互联网地址

如果想浏览或下载本书数据集,请访问 <http://www.stata.com/texts/s4poe4> 或 <http://www.principlesofeconometrics.com>。这里除了可以看到本书其他资料之外,还可以查到单个数据文件。双击你要使用的文件,Stata 会启动并下载之。(当然,你的电脑上得先装好 Stata 软件。)

1.6 变量窗

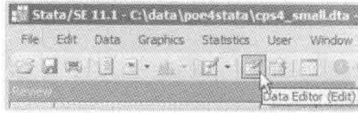
数据文件的变量名列示在 **Variables** 窗口中,同时还列示变量标签(如果有的话)、类型与格式。本书下文的截屏里我们将仅列示变量名和标签。

Name	Label	Type	Format
wage	earnings per hour	double	%10.0g
educ	years of education	byte	%8.0g
exper	post education years experience	byte	%8.0g
hrswk	usual hours worked per week	byte	%8.0g
married	= 1 if married	byte	%8.0g
female	= 1 if female	byte	%8.0g
metro	= 1 if lives in metropolitan area	byte	%8.0g
midwest	= 1 if lives in midwest	byte	%8.0g
south	= 1 if lives in south	byte	%8.0g
west	= 1 if lives in west	byte	%8.0g
black	= 1 if black	byte	%8.0g
asian	= 1 if asian	byte	%8.0g

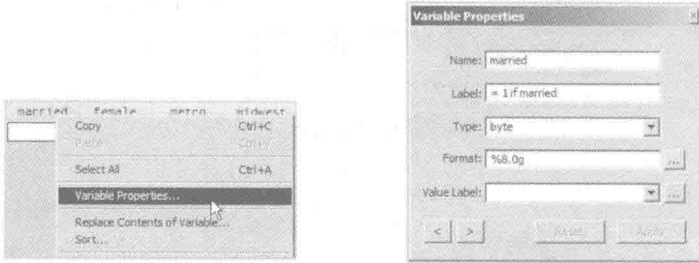
值得注意的是标签很有用,而且很便于添加、修改和删除。

1.6.1 使用数据编辑器修改变量标签

在 Stata 的下拉菜单里选择 **Data Editor** 按钮。



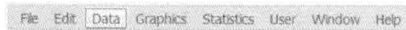
在弹出的结果页面窗里,右击对应的变量栏并选择 **Variable Properties**。



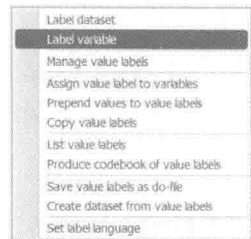
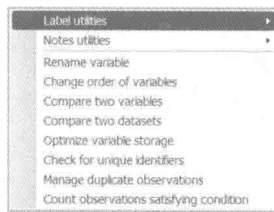
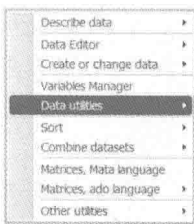
在 **Label** 栏输入修改后的变量标签内容,然后点击 **Apply** 按钮。

1.6.2 使用数据工具修改变量标签

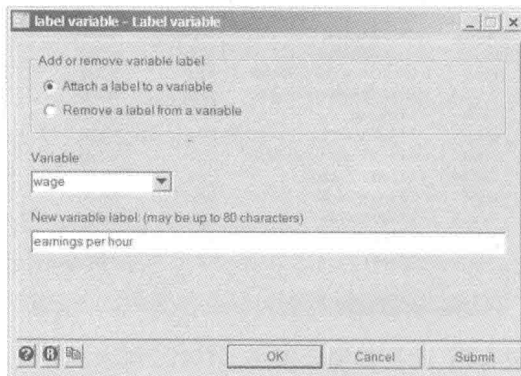
在 Stata 下拉菜单里选择 **Data > Data utilities > Label utilities > Label Variable**,有:



继续:



在最终出现的对话框里,你可以选择 **Attach** a label to a variable,然后在 **Variable** 里通过下拉菜单确定需要改变标签的变量,在 **New variable label** 框里输入新的变量标签,最后点击 **OK** 按钮。



当然你也可以不通过对话框的方式,而是直接在 **Command** 窗口输入:

```
label variable wage "earnings per hour"
```

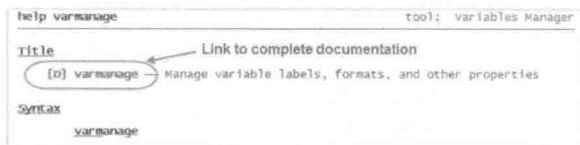
这样就会为 **wage** 变量生成一个新的标签,这个标签同时覆盖了此前的旧标签。当然在对话框中,还可以选择删除标签。

1.6.3 使用变量管理器

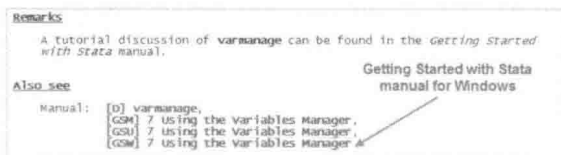
变量管理器能够一站式管理数据变量;在 Stata 下拉菜单里点击 **Variables Manager** 图标即可。



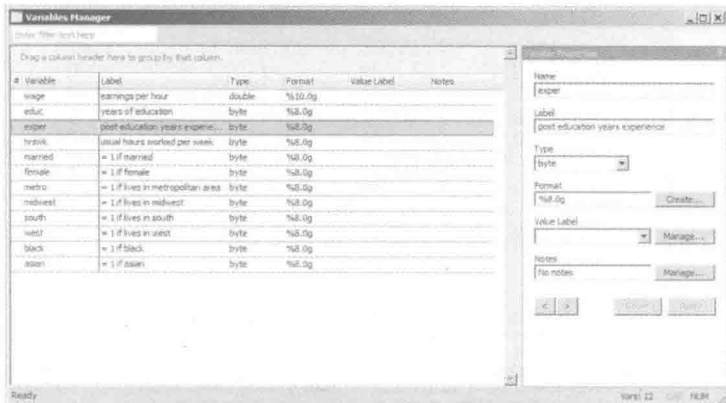
我们还可通过输入命令 `help varmanage` 来获得关于变量管理器特点的更多延伸帮助。在结果观察窗,会有一个关于 `varmanage` 的完整文档链接。



同时还可以看到 `varmanage` 命令的语法,有下划线的部分表示打开变量管理器 **Variables Manager**,只需在命令窗口输入这几个字母。在结果观察窗的最下一行字是关于 **Variables Manager** 的 Stata 的入门指南链接(见下图)。

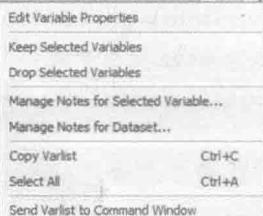


在变量编辑器中,点击某个变量可以打开 **Variable Properties** 窗口,看到该变量的所有特征。此处可以修改变量标签,添加说明,以及单独和批量管理变量。



在变量栏高亮后右击鼠标,可以看到更多的选项。

educ	years of education	byte	%8.0g
exper	post education years experie...	byte	%8.0g
hrswk	usual hours worked per week	byte	%8.0g
married	= 1 if married	byte	%8.0g
female	= 1 if female	byte	%8.0g
metro	= 1 if lives in metropolitan area	byte	%8.0g
midwest	= 1 if lives in midwest	byte	%8.0g
south	= 1 if lives in south	byte	%8.0g
west	= 1 if lives in west	byte	%8.0g
black	= 1 if black	byte	%8.0g
asian	= 1 if asian	byte	%8.0g



1.7 数据描述与概要统计量

每次打开数据文件,需要做一些常规的操作。首先输入命令:

```
describe
```

这个命令能简要介绍内存中的数据集合信息,包括变量名及其标签等。部分结果如下所示:

```
. describe
```

```
Contains data from cps4_small.dta
obs:      1,000
vars:     12
size:     23,000 (99.9% of memory free)
```

variable name	storage type	display format	value label	variable label
wage	double	%10.0g		earnings per hour
educ	byte	%8.0g		years of education
exper	byte	%8.0g		post education years experience
hrswk	byte	%8.0g		usual hours worked per week
married	byte	%8.0g		= 1 if married
female	byte	%8.0g		= 1 if female

接着,输入概要统计命令:

```
summarize
```

在结果窗中会报告概要统计量,部分结果如下所示:

```
. summarize
```

Variable	Obs	Mean	Std. Dev.	Min	Max
wage	1000	20.61566	12.83472	1.97	76.39
educ	1000	13.799	2.711079	0	21
exper	1000	26.508	12.85446	2	65
hrswk	1000	39.952	10.3353	0	90
married	1000	.581	.4936423	0	1
female	1000	.514	.5000541	0	1

当然,也可以通过 Stata 的下拉菜单来实现同样的结果。比如,点击 Stata 面板菜单上的 **Statistics** 按钮图标。



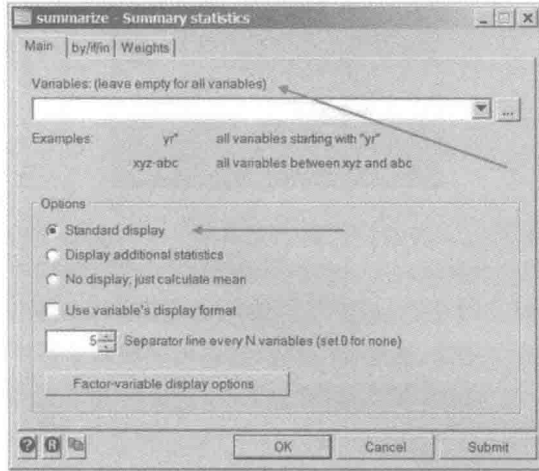
然后会出现一长列各种各样的统计分析工具。如下图所示,现在你应该依次选择点击 **Summaries, tables, and tests**, 然后选择 **Summary and descriptive statistics**, 接着 **Summary statistics**。



完整的点击路径如下：

Statistics > Summaries, tables, and tests > Summary and descriptive statistics > Summary statistics

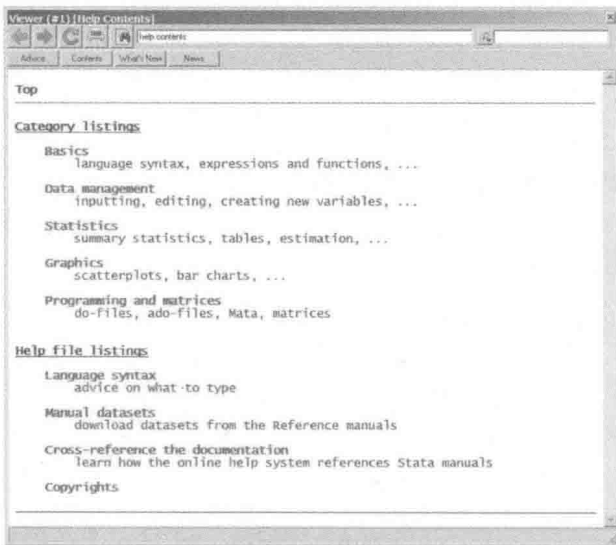
随即弹出的对话框会出现很多选项。基本的概要描述统计表没有备选项，只要点 **OK** 键就可以了，**Stata** 会自动给出数据集中所有变量的概要统计量。也可以在变量栏里输入具体的变量名来选择单一变量。标准的结果输出包括观测样本数、算术平均值、标准差、最小值与最大值。



1.8 Stata 帮助系统

Stata 的帮助系统是它的特色之一。点击 **Stata** 面板栏上的 **Help** 按钮，然后选择 **Contents**。

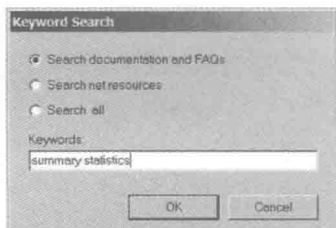
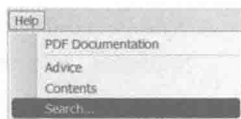
其后出现的窗口中，可以点击蓝色的如 **Basics**, **Data Management** 等分类子项从而转到更进一步的内容。



1.8.1 使用关键词搜索

点击 **Help > Search**:

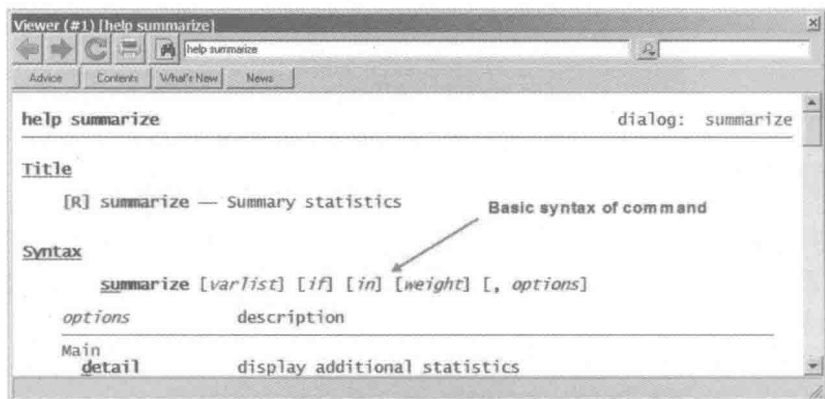
在对话框里会出现多个备选项。在 Stata 文档和 FAQs(常见问题及回答)里搜寻相关资料,只需简单地输入你想要的短语,比如搜索概要统计,则在关键词框中输入 **Summary Statistics**。



如果命令行输入为:

```
search summary statistics
```

随即会出现一系列可能相关的主题,其中蓝色的词汇可以直接链接,比如点击 **Summarize**。结果对话框会直接展示命令语法,这些命令以及选项可以在命令窗中输入。



提示:注意到语法中 su 标了下划线。这是最小缩略形式,意味着 **summarize** 命令可以缩略成 **su**,或者说 **summ**。

更大范围的关键词搜索可以使用 **findit** 命令,例如:

```
findit mixed models
```

更多的搜索选项请使用 **help search**。

1.8.2 使用命令搜索

如果你知道需要了解的 **Stata** 具体命令的名字,你可以点击 **Help > Stata Command**。在弹出的结果对话框中输入具体命令的名字,然后点击 **OK**。