

# Stata软件基本操作和数据分析入门

## 第一讲 Stata操作入门

### 第一节 概况

**Stata** 最初由美国计算机资源中心（**Computer Resource Center**）研制，现在为 **Stata** 公司的产品，其最新版本为 **7.0** 版。它操作灵活、简单、易学易用，是一个非常有特色的统计分析软件，现在已越来越受到人们的重视和欢迎，并且和 **SAS**、**SPSS** 一起，被称为新的三大权威统计软件。

**Stata** 最为突出的特点是短小精悍、功能强大，其最新的 **7.0** 版整个系统只有 **10M** 左右，但已经包含了全部的统计分析、数据管理和绘图等功能，尤其是他的统计分析功能极为全面，比起 **1G** 以上大小的 **SAS** 系统也毫不逊色。另外，由于 **Stata** 在分析时是将数据全部读入内存，在计算全部完成后才和磁盘交换数据，因此运算速度极快。

由于 **Stata** 的用户群始终定位于专业统计分析人员，因此他的操作方式也别具一格，在 **Windows** 席卷天下的时代，他一直坚持使用命令行 / 程序操作方式，拒不推出菜单操作系统。但是，**Stata** 的命令语句极为简洁明快，而且在统计分析命令的设置上又非常有条理，它将相同类型的统计模型均归在同一个命令族下，而不同命令族又可以使用相同功能的选项，这使得用户学习时极易上手。更为令人叹服的是，**Stata** 语句在简洁的同时又拥有着极高的灵活性，用户可以充分发挥自己的聪明才智，熟练应用各种技巧，真正做到随心所欲。

除了操作方式简洁外，**Stata** 的用户接口在其他方面也做得非常简

洁，数据格式简单，分析结果输出简洁明快，易于阅读，这一切都使得 **Stata** 成为非常适合于进行统计教学的统计软件。

**Stata** 的另一个特点是他的许多高级统计模块均是编程人员用其宏语言写成的程序文件（**ADO** 文件），这些文件可以自行修改、添加和下载。用户可随时到 **Stata** 网站寻找并下载最新的升级文件。事实上，**Stata** 的这一特点使得他始终处于统计分析方法发展的最前沿，用户几乎总是能很快找到最新统计算法的 **Stata** 程序版本，而这也使得 **Stata** 自身成了几大统计软件中升级最多、最频繁的一个。

由于以上特点，**Stata** 已经在科研、教育领域得到了广泛应用，**WHO** 的研究人员现在也把 **Stata** 作为主要的统计分析工作软件。

## 第二节 **Stata** 操作入门

### 一、**Stata** 的界面

图 1 即为 **Stata 7.0** 启动后的界面，除了 **Windows** 版本的软件都有的菜单栏、工具栏，状态栏等外，**Stata** 的界面主要是由四个窗口构成，分述如下：

1. **结果窗口**：位于界面右上部，软件运行中的所有信息，如所执行的命令、执行结果和出错信息等均在这里列出。窗口中会使用不同的颜色区分不同的文本，如白色表示命令，红色表示错误信息。

2. **命令窗口**：位于结果窗口下方，相当于 **DOS** 软件中的命令行，此处用于键入需要执行的命令，回车后即开始执行，相应的结果则会在结果窗口中显示出来。

3. **命令回顾窗口**：即 **review** 窗口，位于界面左上方，所有执行

过的命令会依次在该窗口中列出,单击后命令即被自动拷贝到命令窗口中;如果需要重复执行,用鼠标双击相应的命令即可。

4. **变量名窗口**: 位于界面左下方,列出当前数据及中的所有变量名称,。

除以上四个默认打开的窗口外,在 **Stata** 中还有数据编辑窗口、程序文件编辑窗口、帮助窗口、绘图窗口、**Log** 窗口等,如果需要使用,可以用 **Window** 或 **Help** 菜单将其打开。

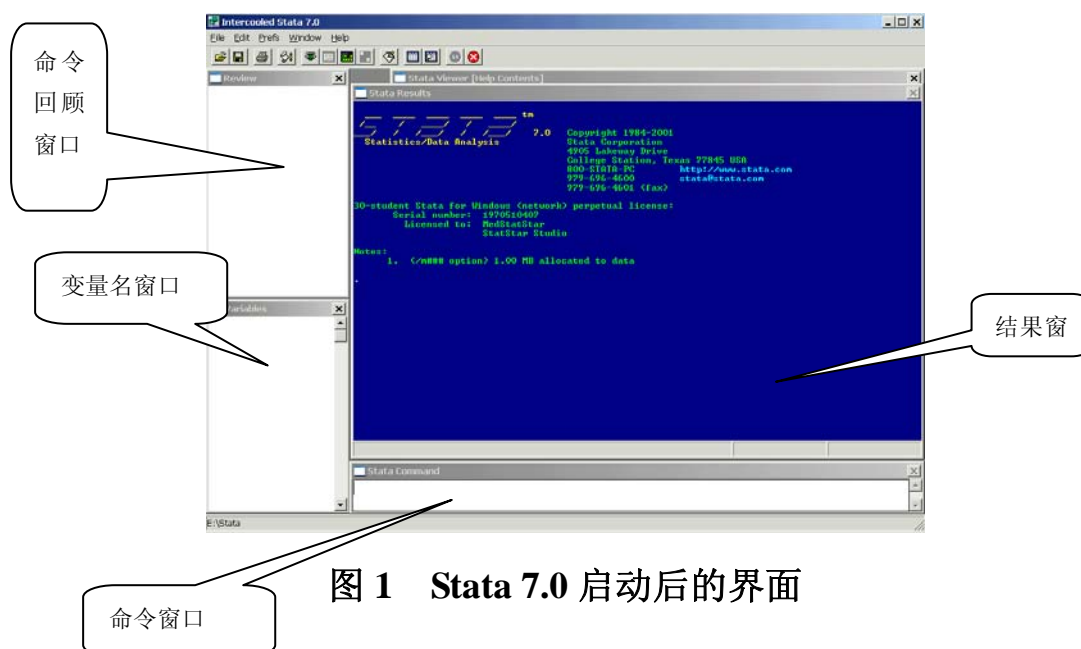


图 1 Stata 7.0 启动后的界面

## 二、数据的录入与储存

**Stata** 为用户提供了简捷,但是非常完善的数据接口,熟悉它的使用方法是使用 **Stata** 的第一步,在 **Stata** 中读入数据可以有三种方式:直接从键盘输入、打开已有数据文件和拷贝、粘贴方式交互数据。

### 1)从键盘输入数据

在 **Stata** 中可以使用命令行方式直接建立数据集,首先使用 **input** 命令制定相应的变量名称,然后一次录入数据,最后使用 **end** 语句表明数据录入结束。

例 1 在某实验中得到如下数据，请在 Stata 中建立数据集。

观测数据

X 1 3 5 7 9

Y 2 4 6 8 10

解：此处需要建立两个变量 X、Y，分别录入相应数值，Stata 中的操作如下，其中划线部分为操作者输入部分。

```
. drop all
```

```
. input x y
```

```
    x    y
```

```
1. 1  2
```

```
2. 3  4
```

```
3. 5  6
```



```
4. 7  8
```

```
5. 9  10
```

```
6. end
```

## 2) 用 stata 的数据编辑工具

### ① 进入数据编辑器

进入 stata 界面，在命令栏键入 edit 或在 stata 的 window 下拉菜单中单击 data editor 或点击编辑图标  (注意： 是浏览图标，点击后只能浏览，不能编辑) 即可进入 stata 数据编辑器。  
(stata 界面如下图 2)



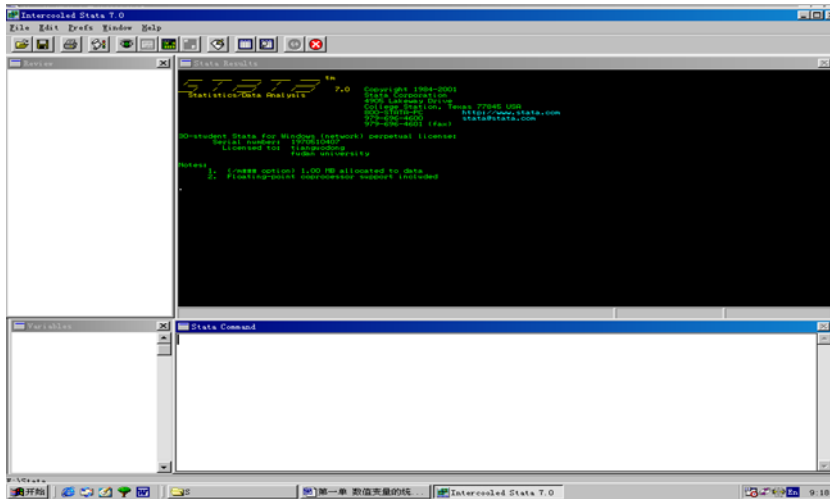


图 2

## ②数据编辑

stata 数据编辑器界面：此时进入了数据全屏幕编辑状态。

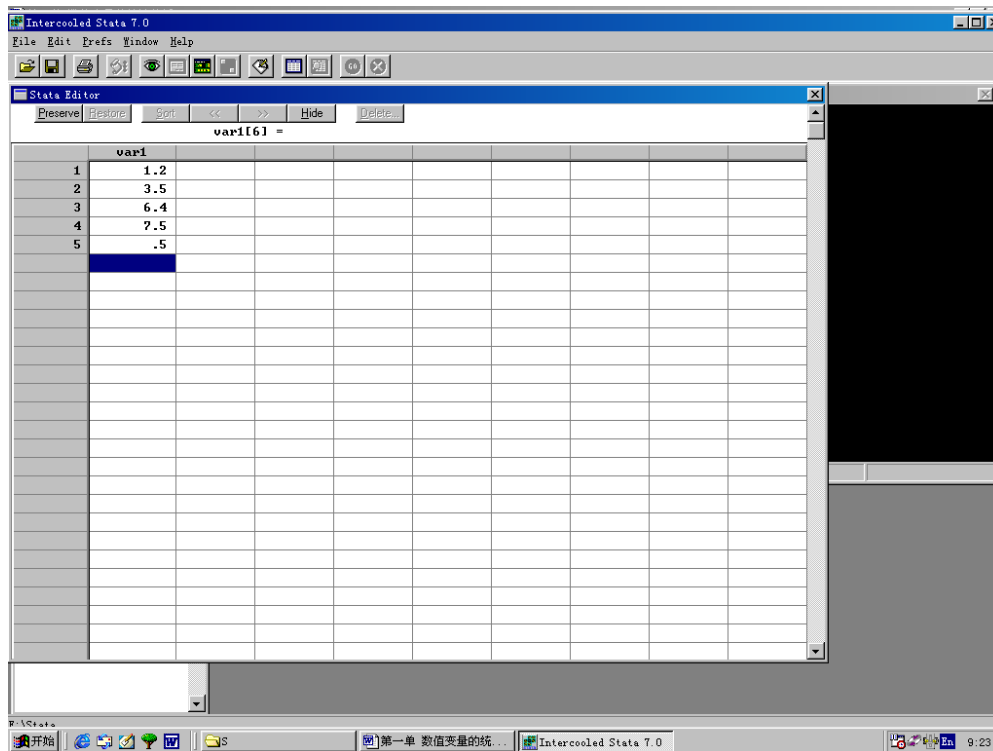


图 3

在第一列输入数据后，Stata 第一列自动命名为 var1；在第二列输入数据后，第二列自动命名为 var2……依次类推。在输入数据后，双击纵格顶端的变量名栏(如：Var1 或 Var2 处)，可以更改变量名，

并可以在 label 栏中注释变量名的含义, 点击 **OK** 确认(如图 4 所示)。

仍沿用上例, 双击观察值所在列顶端的变量名栏, 更改变量为  $x$ , 并在 label 栏中注明 “7 岁男童身高 (cm)”。

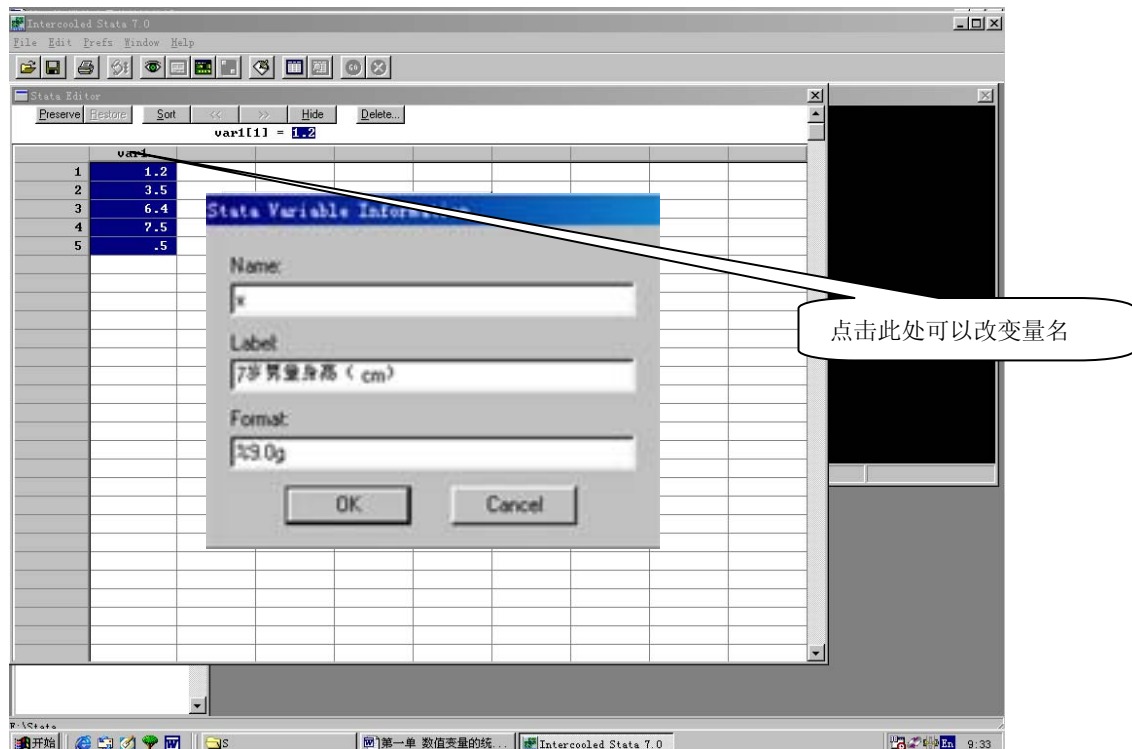


图 4

数据输入完毕后, 单击 **preserve** 键确认所输数据, 按关闭键 **✕** 即可退出编辑器。

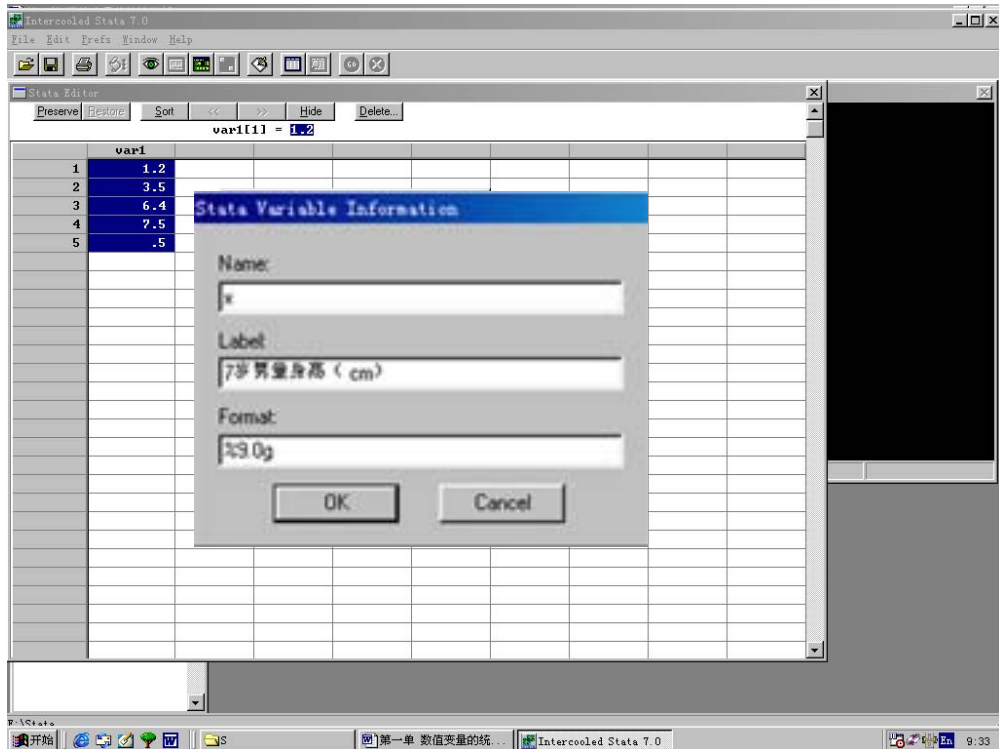


图 5

数据输入完毕后，单击 **preserve** 键确认所输数据，按关闭键  即可退出编辑器。

### 3) 拷贝、粘贴方式交互数据

**Stata** 的数据编辑窗口是一个简单的电子表格，可以使用拷贝、粘贴方式直接和 **EXCEL** 等软件交互数据，在数据量不大时，这种方式操作极为方便。

**例 2** 现在 **EXCEL** 中已录入了三个变量，共五条记录，格式见下图，请将数据读入 **Stata**。

**解：**首先将 **EXCEL** 中的 A1~C6 全部 18 个单元格选中，选择菜单编辑→复制，将数据拷贝到剪贴板上；然后切换到 **Stata**，选择菜单 **Window**→**Data Editor**，打开数据编辑窗口；再选择 **Edit**→**Paste**，相应的数据就会被直接粘贴如数据编辑窗口中，并且变量名、记录数、

变量格式等均会被自动正确设置，见图 6 和图 7。


	A	B	C
1	x	y	z
2		1	2 q
3		3	4 wqw
4		5	6 e
5		7	8 dfw
6		9	10 f

	x	y	z
1	1	2	q
2	3	4	wqw
3	5	6	e
4	7	8	dfw
5	9	10	f

图 6 在 EXCEL 中的数据格式 图 7 粘贴入 Stata 后的数据格式

#### 4)、打开已有的数据文件

**Stata** 能够直接打开的数据文件只能是自身专用格式或者以符号分隔的纯文本格式，后者第一行可以是变量名，分述如下：

1. 点击图标 ，然后选择路径和文件名，可以打开 **Stata** 专用格式的数据文件，并且扩展名为 **.dta**。

2. 打开 **Dta** 数据文件：该格式文件是 **Stata** 的专用格式数据文件，也使用 **use** 命令即可打开，例如要打开数据文件“**C:\data1.dta**”，则命令为：

```
. use c:\data1
```

即扩展名可以省略，如果 **Stata** 中已经修改或者建立了数据集，则需要使用 **clear** 选项清除原有数据，命令为：

```
. use c:\data1 , clear
```

3. 读入文本格式数据：需要使用 **insheet** 命令实现，例如需要读入已建立好的文本格式数据文件“**C:\data1.txt**”，则命令为：

```
. insheet using c:\data1.txt
```

该命令会自动识别第一行是否为变量名，以及变量列间的分隔符是 **tab**、逗号还是其他字符。如果 **Stata** 中已经修改或者建立了数据




集，则需要使用 `clear` 选项清除原有数据，命令为：

```
. insheet using c:\data1.txt , clear
```

## 5)数据文件的保存

为了方便以后重复使用，输入 `Stata` 的数据应存盘。`Stata` 实际上只能将数据存为自身专用的数据格式或者纯文本格式，分述如下：

1. 点击图标，然后选择路径和文件名，点击保存。

2.存为 `dta` 格式：可以直接使用文件菜单，也可以使用 `save` 命令操作，如欲将上面建立的数据文件存入“C:\”中，文件名为 `Data1.dta`，则命令为：

```
. save c:\data1
```

```
file c:\data1.dta saved
```

该指令将在 C 盘根目录建立一个名为“`data1.dta`”的 `Stata` 数据文件，后缀 `dta` 可以在命令中省略，会被自动添加。该文件只能在 `Stata` 中用 `use` 命令打开。如所指定的文件已经存在，则该命令将给出如下信息：`file c:\data1.dta already exists`，告诉用户在该目标盘及子目录中已有相同的文件名存在。如欲覆盖已有文件，则加选择项 `replace`。命令及结果如下：

```
. save c:\data1.dta , replace
```

```
file c:\data1.dta saved
```

2. 存为文本格式：需要使用 `outsheet` 命令实现，该命令的基本格式如下。

```
outsheet [变量名列表] using 文件名 [, nonames replace ]
```

其中变量名列表如果省略，则将全部变量存入指定文件。

如欲将上面建立的数据文件存入文本文件“C:\data1.txt”中，则命令为：

**.outsheet using c:\data1.txt**

此时建立的文件 data1.txt 第一行为变量名，第 2~6 行为变量值。变量列间用 Tab 键分隔。如果不希望在第一行存储变量名，则可以使用 nonames 选项。如果文件已经存在，则需要使用 replace 选项。