

Stata^{*}: 论文中数据分析的一把利剑

或 120 分钟搞定论文数据分析及结果输出

版本号: 1.0

中南财经政法大学 STATA 协会[†] 学术组

2013 年 4 月 18 日

摘要

本文是写给那些没有用过 Stata 软件, 同时又想在论文中有一些数据分析的本科生和研究生的, 使其能在 2 个小时内掌握 Stata 软件最基本的分析功能, 并能将结果按照学术论文的格式输出到 Word 等软件中。本文将一步步地为你展示如何用 Stata 软件来进行基础的数据分析, 并演示生成高质量的描述统计表格、回归结果及统计图并将之在论文中报告出来的详细过程。由于 Stata 软件把 EViews, SPSS 的傻瓜式菜单和 SAS 的命令、编程完美结合起来, 所以它受到了初学者和高级用户的普遍欢迎。Stata 软件体积小巧、简单易懂且功能强大, 相信它会帮助你完成数据分析及报告的任务, 让你得到一篇高水准的论文。

关键词: 描述统计, 计量模型, 回归结果, 数据分析, 实证分析, 论文写作

目录

1	简介	2
1.1	Stata 简介	2
1.2	Stata 软件的安装及界面介绍	2
1.3	Stata 操作方法介绍	2
1.4	Do-file 的使用	4
1.5	中文字符的支持	5
2	准备工作	5
2.1	数据的导入	5
2.2	下载扩展包	7
3	数据分析	7
3.1	描述性统计	8
3.2	相关系数矩阵	8
3.3	最小二乘回归	9
4	结果报告	10
4.1	logout 命令 (统计表、相关系数矩阵表)	10
4.2	esttab 命令 (最小二乘回归表)	11
5	统计制图	12

^{*}本文中所使用的 Stata 软件版本为: Stata/SE 12.0 for Windows (32-bit)

[†]新浪微博名: stataclub。如果对本文有任何改进的意见, 欢迎 @stataclub 或发私信给我们。

1 简介

1.1 Stata 简介

Stata 与 SPSS、SAS 并称为当今三大统计软件。与后者相比，Stata 体积小巧、简单易懂且功能强大。Stata 把 EViews、SPSS 的傻瓜式菜单和 SAS 的命令、编程完美结合起来，所以它一推出就受到了初学者和高级用户的普遍欢迎。Stata 不仅在统计方面功能齐全，其在计量分析领域更是有着深刻影响，以至于有人一言以蔽之：“关于学习 Stata 的意义，大家只需知道：目前，Stata 是计量经济学，特别是微观计量经济学的主流软件。”[1]

1.2 Stata 软件的安装及界面介绍

下载好的 Stata 一般是绿色版压缩包，解压好后出现图 1 的文件夹即可。选中的 StataSE.exe 图标即为 Stata 软件，双击后得到图 2 的界面则说明打开了软件。

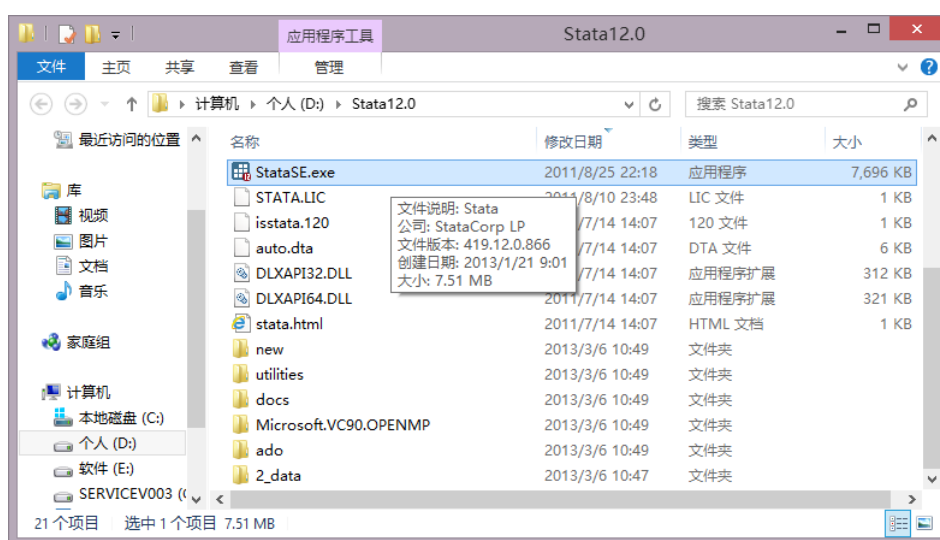


图 1: 解压后的绿色版 Stata 12.0 文件夹

在图 2 中，中间最大一块为“结果显示窗口”，所有（除图片外）的结果都会在此处显示。其下面为“命令窗口”。它很扁的原因是 Stata 的语句一般都不长。左侧的“历史命令窗口”会显示之前的操作（红色命令行则说明操作有误），单击可将命令语句调入命令窗口，双击可直接运行；右侧上方为打开的数据库，两列分别为变量名（Variable）和标签（Label，即用于对变量进行说明），双击即可将变量名导入到命令窗口。右侧下方为变量详细特征。菜单栏中常用的为“Statistics”，即“统计分析”菜单，可做各种统计分析。快捷按钮中前两个为“打开数据”、“保存数据”，第 7 个至第 9 个分别为“打开 Do-file 窗口”、“编辑数据（Edit）”和“浏览数据（Browse）”。

1.3 Stata 操作方法介绍

在 Stata 中既可以使用类似 EViews、SPSS 的傻瓜式菜单的操作方式，也可以使用类似 SAS 的命令、编程的操作方式。以多元线性回归为例（见图 3），Stata 在进行回归时的三种操作方式如下：

1. 菜单式操作，即在菜单栏中选择 Statistics 选项，然后在里面选择相应的统计分析项目。此处做的是多元线性回归，那么在下拉菜单中选 Linear models and related 中的 Linear regression，然后填入相应的因变量 (*price*)、自变量 (*mpg*, *weight*) 即可。故只要知道统计项目的英文名称，那么使用起来还是比较方便的；

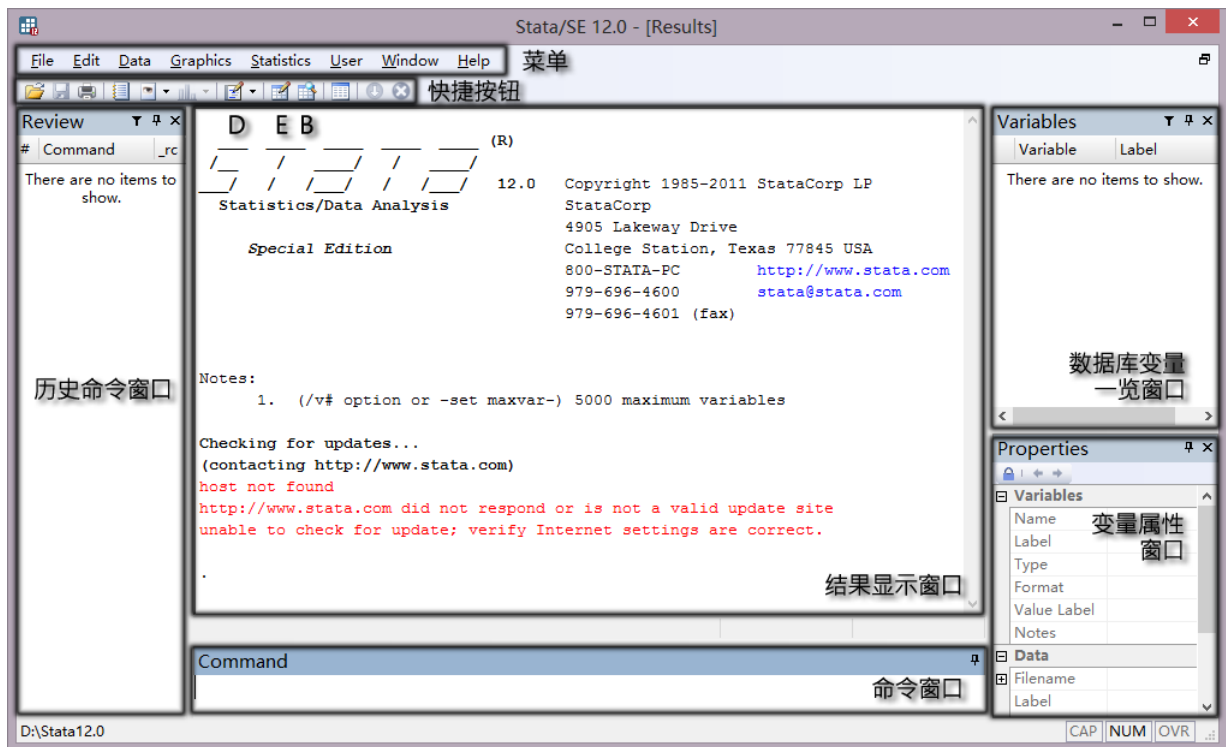


图 2: Stata 12.0 软件界面

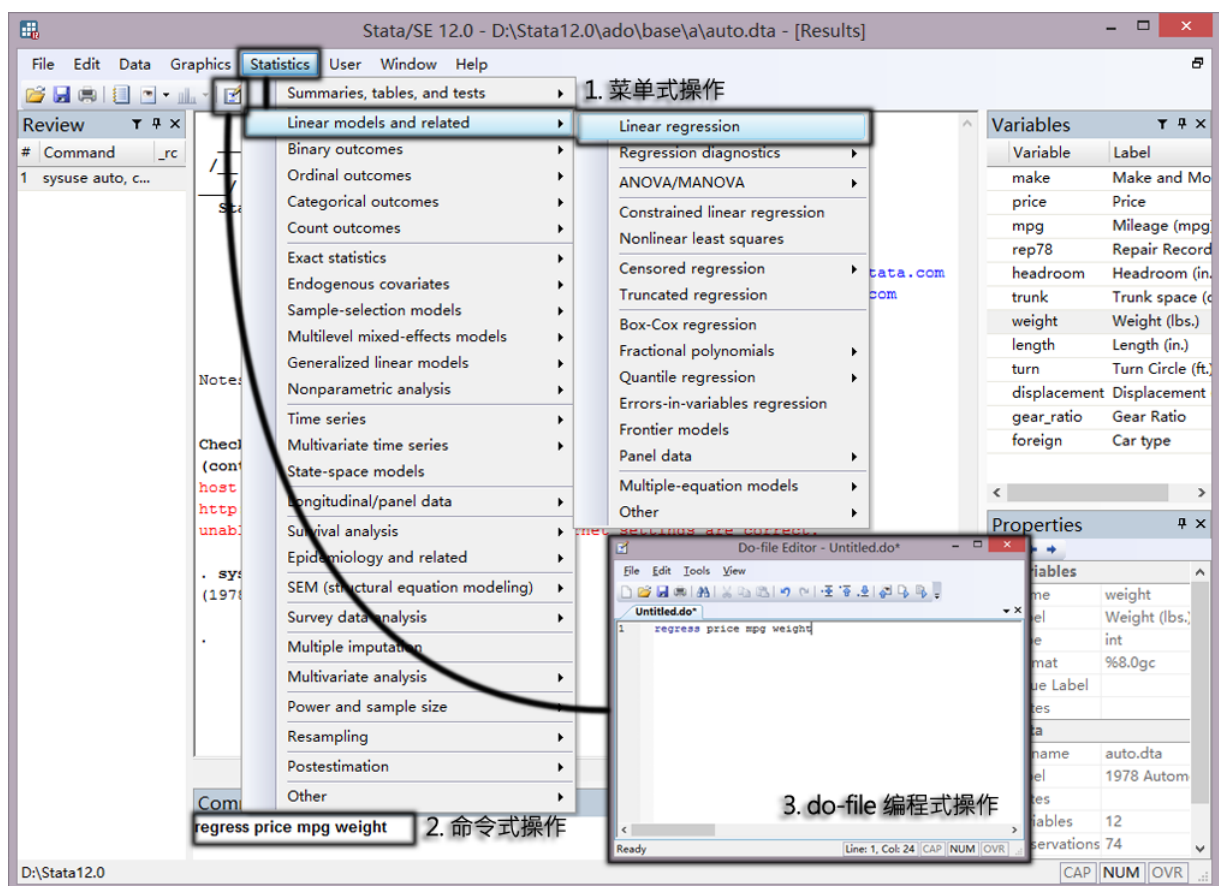


图 3: Stata 中的三种操作方式

2. **命令式操作**，即在 Stata 软件界面下方的 Command 命令窗口直接输入命令进行操作。由于 Stata 的命令简洁方便，所以很多用户在一开始就接受了这种操作方式。同时，在使用菜单式操作时，Stata 会在结果的上方输出对应的命令，所以这也帮助用户熟悉其命令格式。此处对于 $price = \beta_0 + \beta_1 mpg + \beta_2 weight + \varepsilon$ 的回归模型，Stata 命令为：

```
. regress price mpg weight
```

3. **Do-file 程式操作**，即将若干条命令组合在一起，按需求加上一定的循环、条件语句而组成的程序，并运行。程式操作一般是用户对命令非常熟悉时，并要完成批量或更为复杂的任务时所采用的操作方式。但是由于程序有很强的移植性，所以可以从其他人的 Do-file 程序中获取自己想要的，并进行局部修改即可为自己所用，而不必自己重新写代码。同时，Do-file 文档可以保存为 *.do 的文本文件格式，方便修改和以后使用。由于此处只是一句命令语句，并不复杂，所以和命令式操作方式一致。

总结一下：这三种操作方式为初学者和高级用户均提供了合适的环境，一般是逐层递进的。但是我们首推后两种操作方式，因为它们效率高，并有很强的移植性，即简单修改方可为我们所用。而且 Do-file 程式操作方便修改和保存。故我们后续介绍以后两种方式为主。

1.4 Do-file 的使用

由于 Do-file 程式操作的诸多优点，故对其进行详细介绍。在图 2 的快捷按钮中选择第七个（即图中英文字母“D”上面的那个）打开 Do-file 编辑器。我们打开一个 Do-file 窗口，并输入一些 Stata 命令，如图 4 所示。

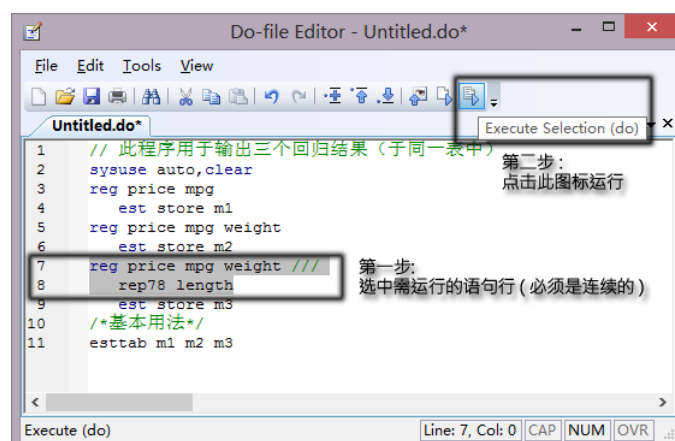


图 4: Stata 的 Do-file 窗口

在图 4 中，如要运行某几行命令语句（必须是连续的），则需要先选中语句，然后点击右上方的“Execute (do)”按钮运行；如果不选中语句而直接点击“Execute (do)”按钮，则会运行此 Do-file 中所有命令语句。注意，运行完 Do-file 后的结果依然是在“结果显示窗口”（见图 2）显示的，故点击按钮后不要傻傻地在 Do-file 窗口等着。

还需要说明的是，程序中绿色部分为注释语句¹，一般用于做标注²。若要注释某行语句可以以“*”、“//”（不包括引号）开始（如图 4 中第 1 行），这样会注释单行；若要注释多行，请将注释内容置于“/*”和“*/”（不包括引号）中（如图 4 中第 10 行）。注意：三个斜杠“///”（不包括引号）并不是注释的意思，而是告诉 Stata 此语句没有写完（如图 4 中第 7 行）。这一般在写长语句时为了层次分明而用。

如果要保存 Do-file 编程结果，可在 Do-file 窗口选择第三个保存命令。保存的文件的扩展名为 *.do，此类文件可在下一次打开 Do-file 窗口后打开，也可用记事本打开。

¹给程序做注释是一个好习惯，因为程序如果写长，其可读性就会变弱。如果在程序命令行后及时做注释，不仅能让程序编写者思路清晰，方便修改、维护及查漏补缺，同时也方便了他人阅读。

²如果你看到某些程序中有些语句被故意地注释掉，那么请不要惊讶为什么这些语句没有被删掉，因为这些语句有可能是程序编写者故意留下来的。其原因可能为：（1）此语句与前后某语句类似，即提供另一种实现方案；（2）此语句与其他语句冲突，但是稍作修改即可使用；（3）程序编写者编写时灵光一现写下的语句，这意味着之后的编写有可能会用得到。

1.5 中文字符的支持

Stata 对中文字符的支持力度不够³，所以在下面的地方需要做适当调整：

- 显示窗口。若中文显示为乱码，请按图 5 的方法修改设置，即在 General Preferences 对话框中将 Color Scheme 改为 Simple 即可；
- Do-file。在 Do-file 窗口中输入、粘贴中文字符没问题。但是若在 Do-file 窗口中修改中文字符，那么一个字符要按两次删除键（“Backspace”键）才能将其删除，否则会出现乱码。这是中文字符在 Stata 中储存方式的原因，初学者不必深究。

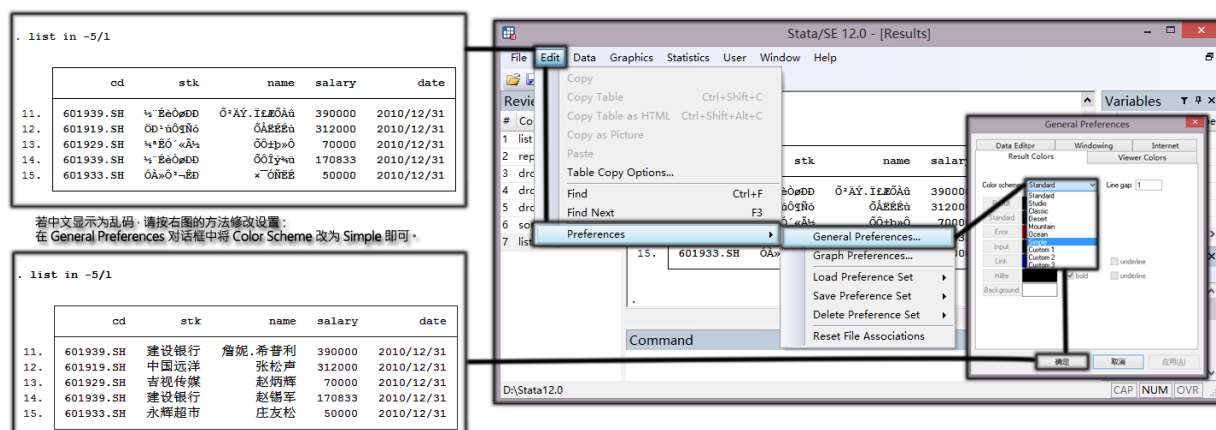


图 5: Stata 显示窗口中中文字符显示乱码的调整办法

2 准备工作

2.1 数据的导入

数据分析的第一步是要有数据。一般我们的数据都是在 Excel 表格中，对于英文界面的 Stata 软件，也许导入数据就够读者头痛的了。下面介绍最简单⁴的导入方法。

Excel 数据 (*.xls/*.xlsx) 此类数据最常见。我们一般操作步骤为：

1. 打开 Excel，如图 6 - (a)。删除第一行的中文变量名，并修改成自定义的英文变量名⁵，如图 6 - (b)；
2. 选中第一行变量名及需要导入的数据（一般“全选”即可），按“复制”；
3. 在 Stata 中按快捷键第 8 个打开数据编辑窗口⁶，右键后按“粘贴”（或快捷键“Ctrl + V”），如图 6 - (c)。此时 Stata 会询问第一行的数据是否视作变量名，点击“Treat first row as variable names”即可自动将第一行英文字符作为变量名并将余下数据导入至 Stata 中，如图 6 - (d)。此时数值变量默认为黑色，字符变量默认为红色。

注意：如果数据类型不统一，或是有非法字符（比如“#”，“=”等），那么导入的时候会出现问题。建议数据在导入 Stata 前需用 Excel 做预处理。

³估计对中国市场重视程度不高，应该今后的版本会改进

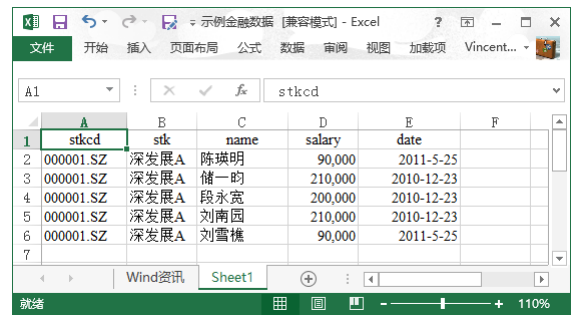
⁴“最简单”指的是在不涉及到批量导入、特殊类型数据导入的情况下导入数据的方法。

⁵变量名不能有空格，且尽可能地简单。一般我们会采用变量英文字母的缩写，如“股票价格 (Stock Code)”缩写为“stkcd”，“收盘价 (Closing Price)”缩写为“clsprc”。

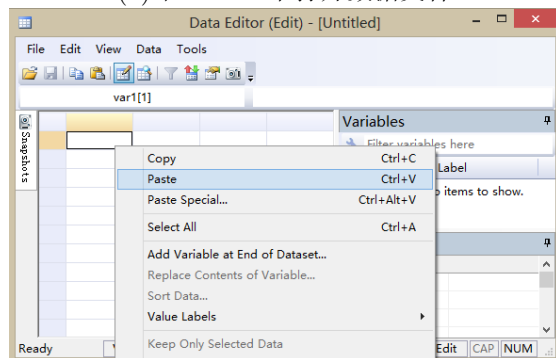
⁶或直接在命令窗口输入 edit 命令并按回车。



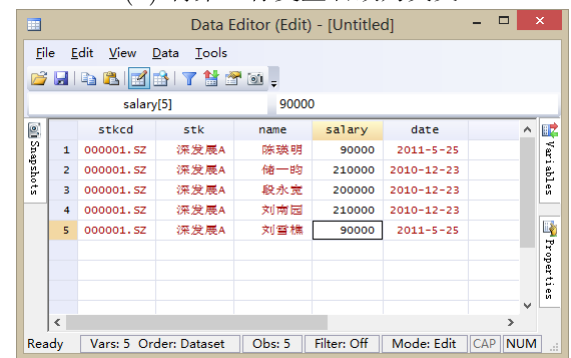
(a) 在 Excel 中打开数据文件



(b) 将第一行变量名改为英文

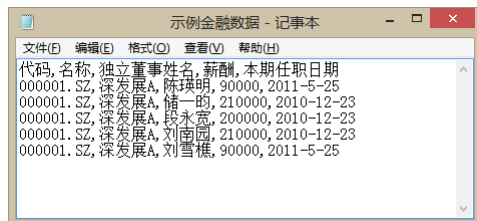


(c) 在 Stata 数据编辑窗口右键按“粘贴”(“Paste”)

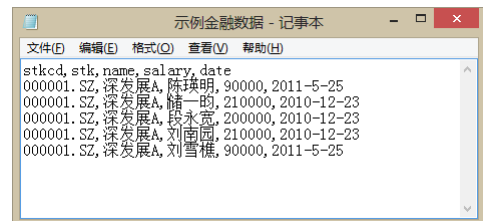


(d) 导入 Stata 后的数据

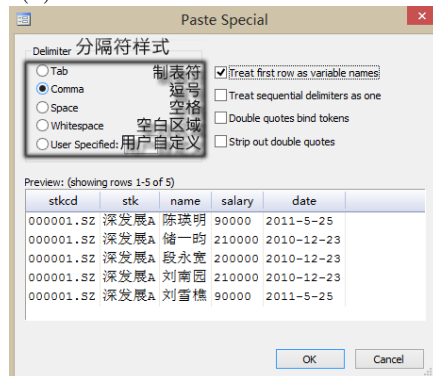
图 6: Excel 型数据的导入



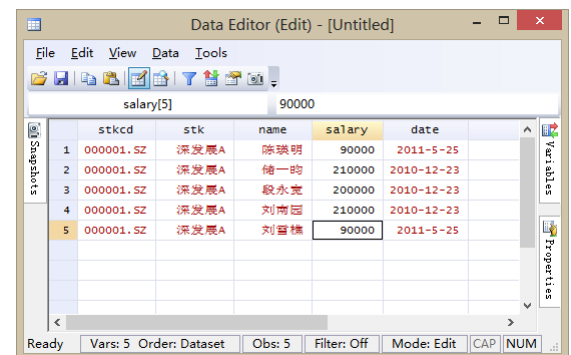
(a) 在记事本中打开 *.csv 数据文件



(b) 将第一行变量名改为英文



(c) 特殊粘贴对话框



(d) 导入 Stata 后的数据

图 7: 逗号分隔型数据的导入

制表符、逗号分隔型数据 (*.txt/*.csv) 及其他类型数据 此类数据导入至 Stata 的方法类似，以逗号分隔型数据 (*.csv) 为例，步骤如下：

1. 用记事本打开 *.csv 数据文件，如图 7 - (a)。删除第一行的中文变量名，并修改成自定义的英文变量名，如图 7 - (b)；
2. 选中第一行变量名及需要导入的数据（一般“全选”即可），按“复制”；
3. 在 Stata 中打开数据编辑窗口，此处右键后选择“特殊粘贴”（或快捷键“Ctrl + Alt + V”），然后弹出对话框如图 7 - (c)。在窗口左侧显示多种数据分隔方式，可分别处理多种分隔格式的数据。如果剪贴板中的源数据是逗号分隔的，那么默认是“逗号 (Comma)”分隔形式。右侧有众多选项，此处我们还是勾选“Treat first row as variable names”，即自动将第一行英文字符作为变量名并将余下数据导入至 Stata 中。点击“OK”即可导入数据，如图 7 - (d)。

数据的打开、保存和清除 当导入上述数据后，数据则保存在内存当中。关闭数据编辑窗口即可在 Stata 的主界面的右上方看到变量名一览表。此时可以进行数据的打开、保存和清除的工作：

- **保存数据**。Stata 内存中的数据是 *.dta 格式，在图 2 所示的主界面点击第二个快捷按钮（或快捷键“Ctrl + S”），指定路径和文件名即可保存数据文件。保存成功后我们发现结果显示窗口多了一行英文命令，而这就是“保存数据”操作的命令形式，其格式为：
`. save [路径]文件名`
- **清除数据**。Stata 在工作时，只能在内存中打开一个 *.dta 文件。如果要打开一个新的数据文件，必须先清空内存中已有的数据文件。执行清除数据可直接在命令窗口输入（输入后敲回车）：
`. clear`
- **打开数据**。在下次开始工作时，想打开已保存的 *.dta 数据文件，可在图 2 所示的主界面点击第一个快捷按钮（或快捷键“Ctrl + O”），找到数据文件打开即可。打开成功后，我们发现结果显示窗口多了一行英文命令，而这就是“打开数据”操作的命令形式，其格式为：
`. use [路径]文件名`

2.2 下载扩展包

Stata 在分析这一块做得近乎完美，但是在结果报告这一块，特别是将结果导入 Word/Excel/LaTeX 等软件时显得不便。好在 Stata 有较好的扩展性，已经有用户编写了扩展命令来解决此类问题，具体使用方法我们会在讲到“结果报告”（见第 4 节）一章再介绍。此处先介绍如何安装这些用户编写的命令。步骤如下：

1. 确保电脑已经联网；
2. 打开 Stata，在命令窗口按照如下格式输入命令：
`. ssc install 命令名 (或者命令组)`
3. 当输入完毕敲击回车后，一般等待半分钟即可下载并安装完毕。在 Stata 结果显示窗口会显示
installation complete.

一般地，我们会用到 `estout`, `outreg2`, `logout` 这几个用于输出结果的扩展包。

3 数据分析

将数据导入 Stata 后，我们便可做统计分析了。Stata 支持的统计分析很多，利用第 1.3 节所介绍的菜单式操作方法几乎可以完成所有的统计分析。不过我们在本文中只介绍最常用的。

3.1 描述性统计

对数据进行最基本的统计运算就是计算出各描述统计值，并将其汇总到一张表中。这里我们用到的是 `tabstat` 命令，其格式如下：

`. tabstat 变量列表, stat(统计量列表)`

例如在读入考试成绩数据文件 `score.dta`（其中包括姓名 `name`，语文成绩 `ch`，数学成绩 `math` 和英语成绩 `eng`）后，我们要列出考试成绩的最高分、最低分、平均分、中位数、标准差和有效试卷数（若要添加其他统计量，请参见表 1），则我们只需键入命令：

`. tabstat ch math eng, stat(max min mean p50 sd n)`

即可得到如下的结果：

				表 1: 常用统计量表			
				统计量	含义	统计量	含义
<code>. tabstat ch math eng, stat(max min mean p50 sd n)</code>				mean	均值	semean	sd/\sqrt{n}
				count	非缺失样本数	skewness	偏度
				n	同上	kurtosis	峰度
				sum	求和	p1	第一分位数
				max	最大值	p5	第五分位数
				min	最小值	p10	第十分位数
				range	max-min	p50	中位数
				sd	标准差	median	同上
				variance	方差	iqr	p75-p25
				cv	sd/mean		

stats	ch	math	eng
max	94	99	94
min	60	50	65
mean	77.43577	74.14541	79.38035
p50	77	73.5	80
sd	10.17742	24.12591	8.156854
N	397	392	390

3.2 相关系数矩阵

若要计算两个变量间的相关系数矩阵，则我们用到的是 `pwcorr` 命令 [2] 来计算变量成对相关系数 (Pairwise Correlation)，其格式如下：

`. pwcorr 变量列表`

如果要顺带要进行 t 检验，且显著性水平 $\alpha = 0.05$ ，那么命令格式为：

`. pwcorr 变量列表, sig star(0.05)`

比如用 Stata 自带的汽车数据 `auto.dta` 来作为例子⁷，现在对于汽车的价格 `price`、行驶里程数 `mpg`、车重 `weight` 以及车身长度 `length` 等变量求相关矩阵，并做 t 检验（显著性水平 $\alpha = 0.05$ ），那么我们依次键入下面两条命令（或编写 Do-file 文件并运行）：

`. pwcorr price mpg weight length`
`. pwcorr price mpg weight length, sig star(0.05)`

有下面的运行结果：

`. pwcorr price mpg weight length`

(Continued on next page)

⁷利用命令 `sysuse auto, clear` 来打开。

	price	mpg	weight	length
price	1.0000			
mpg	-0.4686	1.0000		
weight	0.5386	-0.8072	1.0000	
length	0.4318	-0.7958	0.9460	1.0000

. pwcorr price mpg weight length, sig star(0.05)

	price	mpg	weight	length
price	1.0000			
mpg	-0.4686*	1.0000		
weight	0.5386*	-0.8072*	1.0000	
length	0.4318*	-0.7958*	0.9460*	1.0000

3.3 最小二乘回归

下面介绍经典的**最小二乘法**，也叫做 OLS (Ordinary Least Squares)。如果构造的模型为

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k x_k + \varepsilon$$

那么它在 Stata 中的命令为：

```
. regress y x1 x2 ... xk
```

例如仍旧用上面汽车数据 auto.dta 来作为例子，欲构造的模型为

$$price = \beta_0 + \beta_1 mpg + \beta_2 weight + \beta_3 length + \varepsilon$$

那么应该键入如下命令：

```
. regress price mpg weight length
```

按回车后得到如下结果：

```
. regress price mpg weight length
```

Source	SS	df	MS	Number of obs =	74
Model	226957412	3	75652470.6	F(3, 70) =	12.98
Residual	408107984	70	5830114.06	Prob > F =	0.0000
Total	635065396	73	8699525.97	R-squared =	0.3574
				Adj R-squared =	0.3298
				Root MSE =	2414.6

price	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
mpg	-86.78928	83.94335	-1.03	0.305	-254.209 80.63046
weight	4.364798	1.167455	3.74	0.000	2.036383 6.693213
length	-104.8682	39.72154	-2.64	0.010	-184.0903 -25.64607
_cons	14542.43	5890.632	2.47	0.016	2793.94 26290.93

这个回归报告的左上角是解释/残差/总平方和 (SSE/SSR/SST) 及自由度 (df) 等信息；右上角报告的分别是样本数 (n)、模型 F 检验值、 p 值以及拟合优度 R^2 、调整 R^2 等信息；下半部分则分别列出了各回归变量前的系数 (Coefficient)、标准误 (se)、 t 检验值、 p 值和置信区间的信息。如何对模型进行分析及修正已经超越本文的范畴，相应的 Stata 命令也会在高级的 Stata 书籍内得到详细解释，具体请参见劳伦斯·汉密尔顿 [3] 和陈强 [4] 的有关书籍。

4 结果报告

我们知道，在一篇实证论文中，统计结果的报告与分析是论文中很重要的一块。“实证分析中最重要的就是各种表格”[2] 说的就是这个道理。虽然 Stata 报告的结果简单明了，但是直接粘贴到论文编辑软件中就会显得不是太正式，尤其在回归结果中很多值是冗余的。本节主要介绍如何将 Stata 结果以正规论文的格式来导入 Word 中，至于导入 L^AT_EX 中请主要参照 [5]。

4.1 logout 命令 (统计表、相关系数矩阵表)

logout 命令用于将 Stata 的结果以表格的形式输出到 Word/Excel/L^AT_EX 中。这个命令非常强大，但凡含有表格输出的 Stata 命令，均可以在命令前添加 logout，以使结果转换成需要的格式。这个命令在第一次使用时需要联网下载，具体请参见 2.2 节内容。

logout 命令的格式为：

. logout, 选项 : 主命令

其中选项有输出文件名 save()、输出格式 word, excel, tex 和是否覆盖同名文件 replace。如果输出 test.rtf 文件⁸，那么选项为 save(test) word；由于输出报告结果不是一次就成功的，我们就会反复进行尝试⁹，那么一般会加上 replace 选项；上述选项可以多个使用，使用的时候只需用空格隔开。而 logout 命令格式中的“主命令”则为 Stata 真正需执行的命令。一般我们将某条命令执行成功后，加上前面的 logout 命令及选项即可输出成表格。

例如要将 3.1 和 3.2 节的结果输出至 Excel 文档，那么可执行下面的 Do-file：

```
1 use score, clear
2 logout, save(test1) excel replace: ///
3     tabstat ch math eng, stat(max min mean p50 sd n)
4 sysuse auto, clear
5 logout, save(test2) excel replace: ///
6     pwcorr price mpg weight length, sig star(0.05)
```

然后运行后 Stata 的结果窗口会出现如下蓝色的结果：

```
test2.xml
dir
```

其中第一个就是输出的结果，点击后可以用 Excel 软件打开；第二个点击后会弹出文件所在的文件夹。

需要注意的是，logout 命令得到的表格一般会存在错位和空行的现象，见表 2。这就需要手动做一些调整，如将第一行的表头往右移一列， p 值向右移，并将多余的行删掉，同时还要对 p 值加上括号，并在表的底端注明星号的含义。见表 3。

⁸logout 命令将输出的 Word/Excel 文件扩展名默认为 *.rtf/*.xml 格式，这是因为较 *.doc/*.xls 的常规格式来说，前者不仅可用 Word/Excel 打开，而且还可用其他软件打开，其应用更广。

⁹在反复尝试的时候，记得在运行命令之前将旧的 *.rtf/*.xml 文档关闭，否则新的文档将无法生成，并提示 file *.rtf is read-only; cannot be modified or erased 的错误。

表 2: 输出的相关矩阵表格 (test2.xml)

price	mpg	weight	length	
price	1			
mpg	-0.4686*	1		
0				
weight	0.5386*	-0.8072*	1	
0	0			
length	0.4318*	-0.7958*	0.9460*	1
0.000100	0	0		

表 3: 经调整后的相关矩阵表格

	price	mpg	weight	length
price	1			
mpg	-0.4686* (0.0000)	1		
weight	0.5386* (0.0000)	-0.8072* (0.0000)	1	
length	0.4318* (0.0001)	-0.7958* (0.0000)	0.9460* (0.0000)	1

* 表示通过显著性水平 $\alpha = 0.05$ 的 t 检验。

4.2 esttab 命令（最小二乘回归表）

首先要明确的是，即使 `logout` 命令可将 3.3 节中的回归结果以表格形式输出到 Word 中，我们也千万不要这么做——我们只需要在论文中罗列出回归后的核心结果，而非所有结果。这些结果包括：变量前的系数（必须）、 t 检验值或者标准误（必须）、样本数 n 、模型的拟合优度 R^2 、 F 检验值等。

例如 3.3 节中的回归结果可以下面回归方程式或表格（见表 4）的形式报告出来（括号中的值为标准误）：

$$\text{price} = 14542.43 - 86.79 \text{mpg} + 4.36 \text{weight} - 104.87 \text{length} + \varepsilon$$

(5890.63) (83.94) (1.17) (39.72)

$$n = 74, \quad R^2 = 0.3574, \quad \bar{R}^2 = 0.3298$$

表 4: 表格式回归结果报告

	(1)
	price
mpg	-86.79
	(83.94)
weight	4.365***
	(1.167)
length	-104.9*
	(39.72)
__cons	14542.4*
	(5890.6)
N	74
R^2	0.357
adj. R^2	0.330

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

表 5: 多个回归结果对比

	(1)	(2)
	price	price
mpg	-238.9***	-86.79
	(53.08)	(83.94)
weight		4.365***
		(1.167)
length		-104.9*
		(39.72)
__cons	11253.1***	14542.4*
	(1170.8)	(5890.6)
N	74	74
R^2	0.220	0.357
adj. R^2	0.209	0.330

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

其中第一个回归方程式是无法用 Stata 来生成的，而第二个却可以使用 `esttab` 命令直接生成。而且在多个回归方程做对比的情况下，后者要比前者更有针对性。这个命令在第一次使用时需要联网下载¹⁰，具体请参见 2.2 节内容。

在使用 `esttab` 命令输出之前，我们需要现在 Stata 中运行 `regress` 的回归。接着，输出到 Word 中¹¹ 的 `esttab` 命令格式为：

`. esttab using 文件名.rtf, r2 ar2 se replace nogap`

¹⁰ 此处联网下载的是名为 `estout` 的命令包，但是我们经常使用的是这个包里面的 `esttab` 命令。

¹¹ 输出到 L^AT_EX 中只需将 `esttab` 命令格式中的 *.rtf 改为 *.tex。更多其他格式请在 Stata 命令窗口输入 `help esttab` 来获取帮助。

其中 `r2`, `ar2` 是分别输出 R^2 和 \bar{R}^2 的值; `se` 是表明括号中的值为标准误, 否则会默认输出 t 检验值; `replace` 是覆盖同名文件; `nogap` 是忽略表格中的空行, 否则在默认情况下表格中每个系数下都会有空行, 会显得不够紧凑。

如果要进行多个回归模型的对比, 可以参考下面的 Do-file 文件。结果见表 5。

```
1 sysuse auto, clear
2 regress price mpg
3 est store m1          // 将第一次回归储存为 m1
4 regress price mpg weight length
5 est store m2          // 将第二次回归储存为 m2
6 esttab m1 m2 using reg1.rtf, r2 ar2 se replace nogap
```

5 统计制图

如果要用统计图来描述结果, 首先要明确你的目的, 然后再在众多统计图选择一个。一般统计图有下列几种 (括号中的为特点): **散点图 (观察两变量之间关系)**、折线图 (观察走势)、条形图 (做对比)、饼图 (所占比例)、直方图 (分布)、箱线图 (看五值信息) 等, 这些图的绘制请参见 [6]。在本节中, 我们主要介绍与 3.2 和 3.3 节相关的统计图: 矩阵图和双变量散点 - 回归直线图。如果以 `auto.dta` 数据文件为例, 我们可以写出如下 Do-file。结果见图 8 和图 9。

```
1 sysuse auto, clear
2 graph matrix price length weight mpg          // 画出 price length weight mpg 变量的矩阵图
3 scatter price weight || lfit price weight      // 画出 price, weight 变量的散点图并拟合直线
```

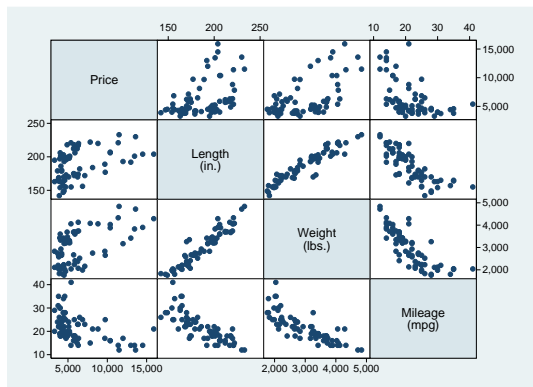


图 8: 矩阵图

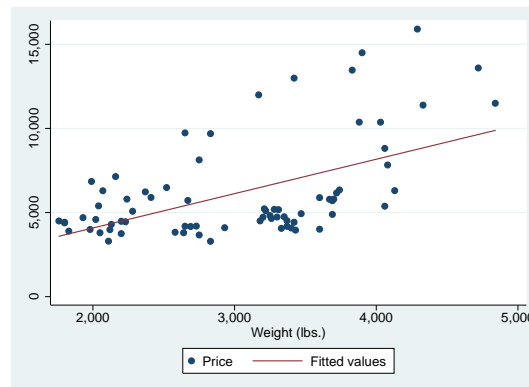


图 9: 散点 - 回归直线图

参考文献

- [1] 《Stata 简明讲义》, 王非, 中国经济研究中心, ebwf@163.com
- [2] 《学会“懒惰”用好 Stata》, 徐鑫, 中南财经政法大学金融学院, 2013.03, xuxinfinance@gmail.com
- [3] 《应用 Stata 做统计分析》, 劳伦斯·汉密尔顿等, 重庆大学出版社, 2011,
- [4] 《高级计量经济学及 Stata 应用》, 陈强, 高等教育出版社, 2010,
- [5] 《Stata 与 L^AT_EX 的完美结合》, 连玉君, 中山大学岭南学院, 2007.08, arlionn@163.com
- [6] 《Stata 画图专题 (2): 基础绘图命令》, 中南财经政法大学 STATA 协会学术组, 中南财经政法大学, 2013.03