

## 第 4 章 Stata 简介

### 4.1 为什么使用 **Stata**

Stata 是目前在欧美最流行的计量软件, 操作简单、功能强大。

使用 Stata 的用户很多, 对于最新计量方法, 常可下载由用户写的 Stata 命令程序(user-written Stata commands), 十分方便。

官方的 Stata 版本也经常更新, 以适应计量经济学的迅猛发展。

Stata 13 已于 2013 年 6 月发布, 但由于在中国普遍使用的仍是 Stata 12 或更低版本, 故本书主要介绍 Stata 12。

## 4.2 Stata 的窗口

安装 Stata 后，点击电脑桌面上的 Stata 图标，即可打开 Stata。

此时可以看到，在最上方有一排菜单，即“File Edit Data Graphics Statistics User Window Help”。

在菜单之下，则为一系列图标，起着快捷键的作用。

在图标之下，有五个窗口，分别为(如图 4.1)

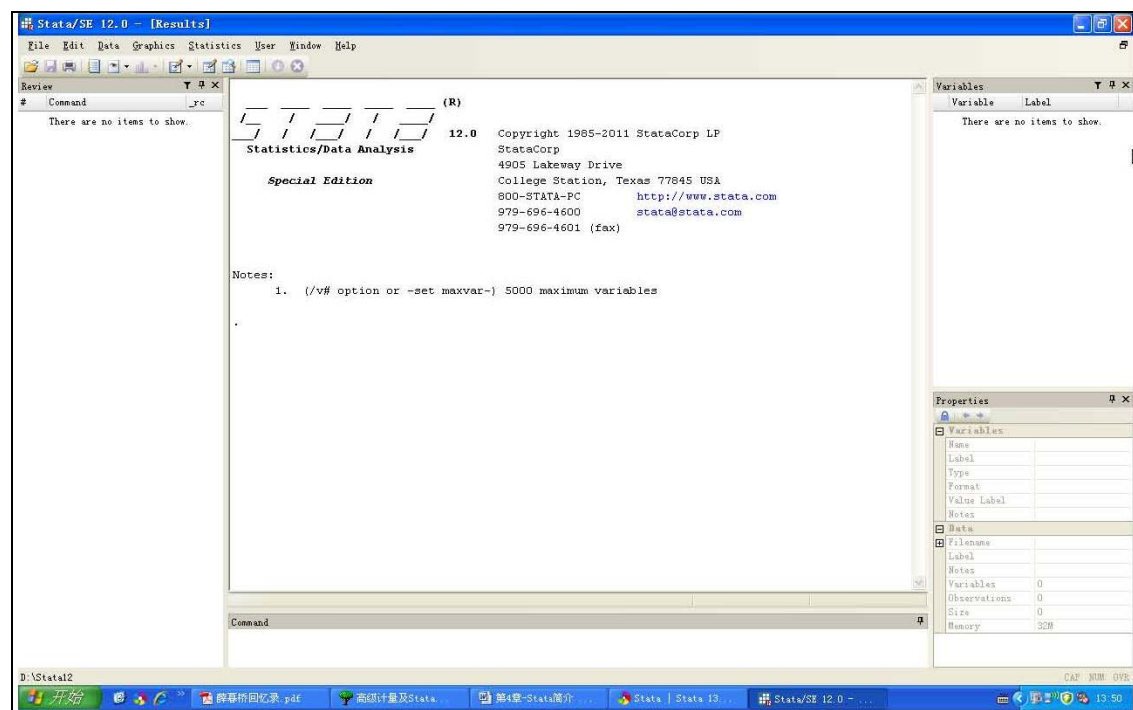


图 4.1 Stata 12 的主要窗口

- 左上 “Review” (历史窗口): 记录着自启动 Stata 以来的命令。
- 中上 “Results” (结果窗口): 显示执行 Stata 命令后的输出结果。
- 中下 “Command” (命令窗口): 在此窗口输入 Stata 命令。
- 右上 “Variables” (变量窗口): 记录着 Stata 内存的所有变量。
- 右下 “Properties” (性质窗口): 显示当前数据文件与变量的性质。

为了使屏幕分割更美观实用，可用鼠标将以上窗口拉到任意大小与位置。

然后点击菜单“Edit”→“Preferences”→“General Preferences”→“Windowing”→“Lock splitter”，锁定当前画面。

在以后重启 Stata 时，将自动显示这个画面设置。

### 4.3 Stata 操作实例

以 Nerlove(1963)对电力行业规模报酬的经典研究来介绍 Stata 的实际操作。该数据集 nerlove.xls(Excel 文件)包括了 1955 年美国 145 家电力企业的横截面数据。

#### 1. 将数据导入 Stata

打开 Stata 软件后, 点击 Data Editor (Edit)图标 (也可点击菜单 “Window” → “Data Editor” ), 即可打开类似 Excel 的空白表格。

用 Excel 打开文件 “nerlove.xls”, 复制所有数据, 粘贴到 Data Editor 中。

Stata 会问你“第一行为数据还是变量名”(Is the first row data or variable names?), 点击相应的选择即可(对于此数据集, 应选“Treat first row as variable names” )。

导入数据的另一方法是(特别在数据量很大的情况下), 点击菜单“File”→“Import”, 然后导入各种格式的数据。但不如直接从 Excel 表中粘贴数据方便直观。

关闭 Data Editor (Edit)后, 即会看到右上方的“Variables”窗口出现了 5 个变量, 分别为 tc(total cost, 总成本), q(total output, 总产量), pl(price of labor, 小时工资率), pf(price of fuel, 燃料价格), 与 pk(user cost of capital, 资本的租赁价格)。

点击 **Save** 图标 (也可点击菜单 “**File**” → “**Save**” ), 将数据存为 Stata 格式的文件(扩展名为 dta), 比如 nerlove.dta。以后就可用 Stata 直接打开此数据集。

打开的方式有两种。方法一, 点击 **Open** 图标 (也可点击菜单 “**File**” → “**Open**” ), 寻找要打开的 dta 文件位置。

方法二, 在命令窗口输入以下命令(假设文件在 E 盘的根目录)并回车(按 **Enter** 键):

```
. use E:\nerlove.dta,clear
```

其中, 逗号 “,” 之后的 “clear” 为 “选择项” (option), 表示可替代内存中的已有数据。

如要关闭一个数据集, 以便使用另外一个数据集, 可输入命令

. clear

内存中数据将被清空，可再打开另外一个数据集。

## 2. 日期数据的导入(可暂时跳过此部分)

## 3. 变量的标签

在变量窗口，变量的“名字”(Name)旁边显示其“标签”(label)。

目前的标签过于简略，缺乏变量的解释信息。

点击倒数第 3 个图标，即可打开变量管理器(Variables Manager)(或点击菜单“Data”→“Variables Manager”)，然后编辑变量名、



标签以及变量的存储格式。

例：把 tc, q, pl, pf 与 pk 的标签分别改为 “total cost”, “total output”, “price of labor”, “price of fuel” 与 “user cost of capital”。

Stata 严格区分大小写字母(case sensitive), 建议对于变量名使用小写字母。

#### 4. 审视数据

想看数据集中的变量名单、标签等, 可输入命令

`. describe`

其中, “describe” 的下划线表示, 可将该命令简写为 “d”。

给数据集加一个标签，说明来自 “Nerlove 1963 paper”:

```
. label data "Nerlove 1963 paper"
```

再次运行命令 “describe”，就会看到数据集的标签 “Nerlove 1963 paper”。

Contains data				
obs:	145	Nerlove 1963 paper		
vars:	5			
size:	2,320			
	storage	display	value	
variable name	type	format	label	variable label
tc	float	%8.0g		total cost
q	int	%8.0g		total output
pl	float	%8.0g		price of labor
pf	float	%8.0g		price of fuel
pk	int	%8.0g		user cost of capital
Sorted by:				
Note: dataset has changed since last saved				

如果想看变量 tc 与 q 的具体数据，可使用命令：

```
. list tc q
```

如想中途停止该命令的执行，可点击 **Break** 图标，或直接在键盘上同时按 “**Ctrl + Break**”。

	tc	q
1.	.082	2
2.	.661	3
3.	.99	4
4.	.315	4
5.	.197	5
6.	.098	9
7.	.949	11
8.	.675	13
9.	.525	13
10.	.501	22
11.	1.194	25
12.	.67	25
13.	.349	35
14.	.423	39
15.	.501	43
16.	.55	63
—Break—		
<u>r(1);</u>		

如改变主意，仍希望显示变量 `tc` 与 `q` 的全部数据：

把光标放在命令窗口，并按键盘上的“Page Up”键即可调用上一命令

使用“Page Down”键可调用下一命令。

另一简便方法是，在左上角的历史窗口点击任何曾用过的命令：

如果用鼠标单击旧命令，则会把旧命令重新调入命令窗口，按回车后即执行，或将旧命令进行编辑后再执行；

如果用鼠标双击旧命令，则将马上自动执行。

只对数据集的一部分执行命令,比如只看 tc 与 q 的前 5 个数据:

```
. list tc q in 1/5
```

	tc	q
1.	.082	2
2.	.661	3
3.	.99	4
4.	.315	4
5.	.197	5

如要罗列从第 32-36 个观测值, 可输入命令:

```
. list tc q in 32/36
```

	tc	q
32.	3.154	214
33.	2.599	220
34.	3.298	234
35.	2.441	235
36.	2.031	253

也可通过逻辑关系来定义数据集的子集。如要列出所有满足条件“ $q \geq 10000$ ”的变量 tc 与 q 的数据，可使用以下命令

```
. list tc q if q>=10000
```

	tc	q
142.	67.12	11477
143.	73.05	11796
144.	139.422	14359
145.	119.939	16719

其中，“>=”表示“大于等于”。其他表示关系的逻辑符号为“=”(等于)，“>”(大于)，“<”(小于)，“<=”(小于等于)，“~=”(不等于)。

查看具体数据的直接方法是，点击 Data Editor (Edit)图标，或者点击该图标右边的 Data Editor (Browse)图标。

如要删除满足 “ $q \geq 10000$ ” 条件的观测值，输入命令

```
. drop if q>=10000
```

如只想保留满足 “ $q \geq 10000$ ” 条件的观测值，可使用命令

```
. keep if q>=10000
```



## 5. 考察变量的统计特征

如果看变量  $q$  的统计特征，可输入命令

```
. summarize q
```

Variable	Obs	Mean	Std. Dev.	Min	Max
q	145	2133.083	2931.942	2	16719

显示变量  $q$  的样本容量、平均值、标准差、最小值与最大值。

如计算满足条件 “ $q \geq 10000$ ” 的子样本的统计指标，使用命令

```
. su q if q >= 10000
```

Variable	Obs	Mean	Std. Dev.	Min	Max
q	4	13587.75	2453.921	11477	16719

如想看更多的统计指标，使用命令

```
. su q,detail
```

total output				
Percentiles		Smallest		
1%	3	2		
5%	13	3		
10%	43	4	Obs	145
25%	279	4	Sum of Wgt.	145
50%	1109	Largest	Mean	2133.083
			Std. Dev.	2931.942
75%	2507	11477		
90%	5819	11796	Variance	8596285
95%	8642	14359	Skewness	2.398202
99%	14359	16719	Kurtosis	9.474916

新增的统计指标有百分位数(percentiles)，方差(variance)，偏度(skewness)与峰度(kurtosis)。

如果不指明变量，将显示数据集中所有变量的统计指标。

```
. su
```

Variable	Obs	Mean	Std. Dev.	Min	Max
tc	145	12.9761	19.79458	.082	139.422
q	145	2133.083	2931.942	2	16719
pl	145	1.976552	.2300404	1.5	2.3
pf	145	26.17655	7.876071	10.3	42.8
pk	145	174.4966	18.20948	138	233

如果要显示变量 `pl` 的经验累积分布函数(empirical cumulative distribution function)，可使用命令

```
. tabulate pl
```

price of labor	Freq.	Percent	Cum.
1.5	7	4.83	4.83
1.6	4	2.76	7.59
1.7	15	10.34	17.93
1.8	26	17.93	35.86
1.9	12	8.28	44.14
2	12	8.28	52.41
2.1	32	22.07	74.48
2.2	17	11.72	86.21
2.3	20	13.79	100.00
Total	145	100.00	

如要显示内存中 3 个价格变量之间的相关系数，输入命令

```
. pwcorr pl pf pk,sig star(.05)
```

选择项 “sig” 表示显示相关系数的显著性水平(即  $p$  值，列在相关系数的下方)，选择项 “star(.05)” 表示给所有显著性水平小于或等于 5% 的相关系数打上星号。

如果 `pwcorr` 之后没有指定变量, 则显示所有变量的相关系数。

	p1	pf	pk
p1	1.0000		
pf	0.3310* 0.0000	1.0000	
pk	-0.1845* 0.0263	0.1254 0.1328	1.0000

pf 与 p1 的相关系数为 0.331, 在 5% 水平上显著( $p$  值为 0.0000);

pk 与 p1 的相关系数为-0.1845, 在 5% 水平上显著( $p$  值为 0.0263);

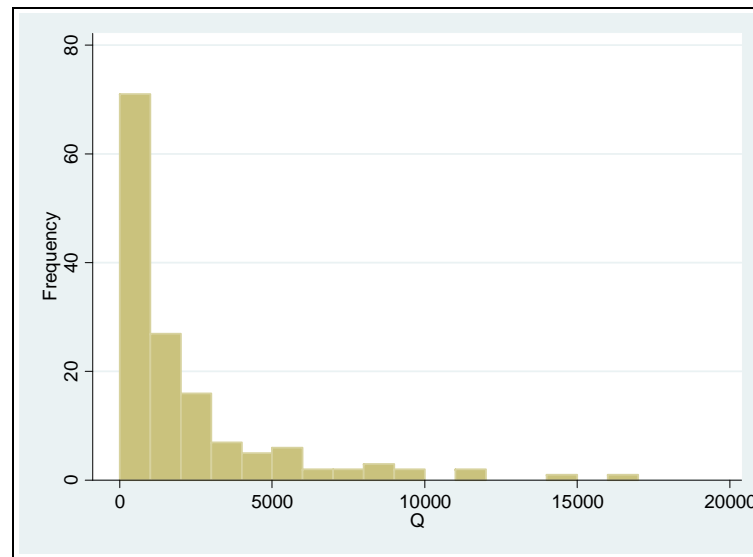
pk 与 pf 的相关系数为 0.1254, 在 5% 水平上不显著( $p$  值为 0.1328)。

## 6. 画图

画变量  $q$  的直方图(假定组宽为 1000), 输入命令

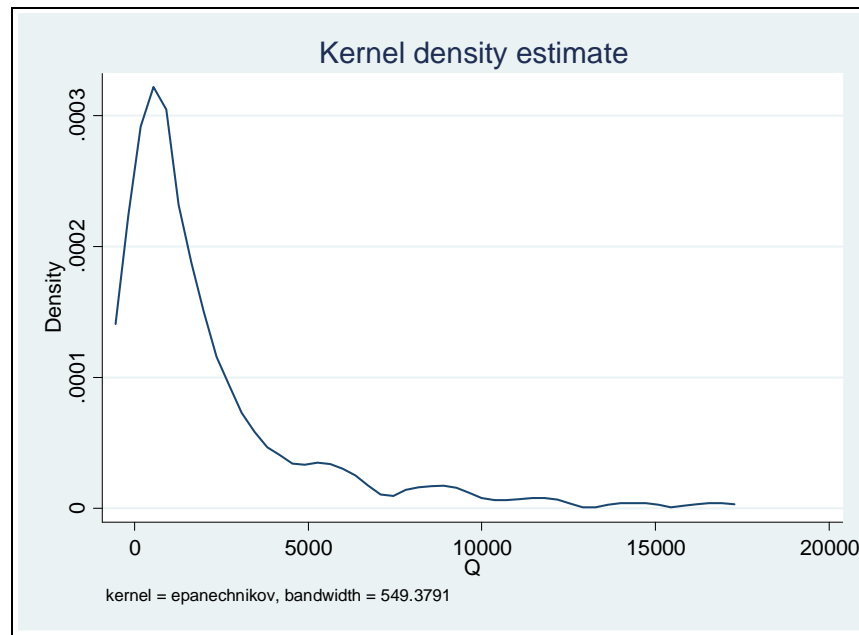
```
. histogram q, width(1000) frequency
```

逗号 “,” 之后的 “width(1000)” 与 “frequency” 为 “选项” (options), 分别表示将组宽设为 1000, 将纵坐标定为频数。



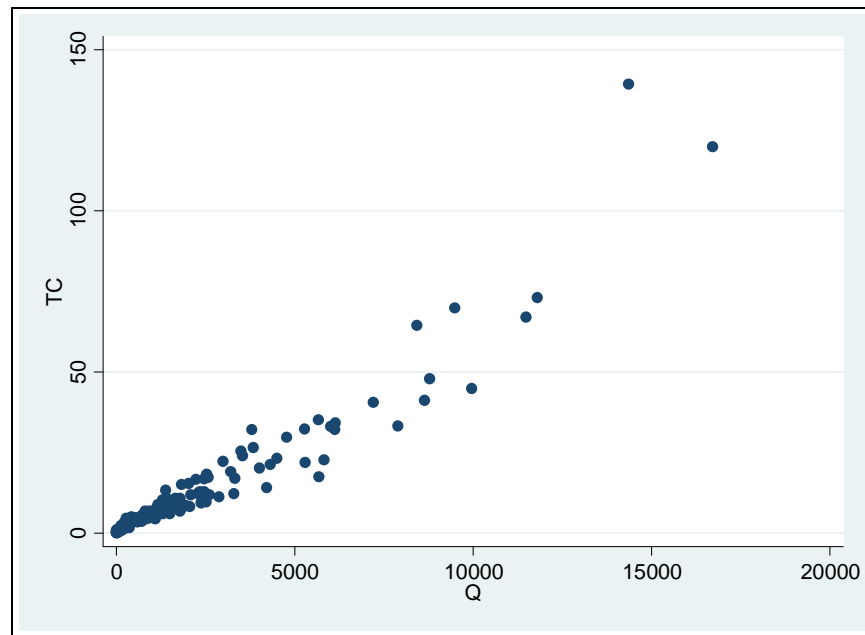
直方图不连续。如看连续的经验分布图（核密度图），使用命令：

```
. kdensity q
```



如画 tc 与 q 之间的散点图，输入命令：

```
. scatter tc q
```





在上页的散点图中，无法知道每个点分别对应哪个观测值。

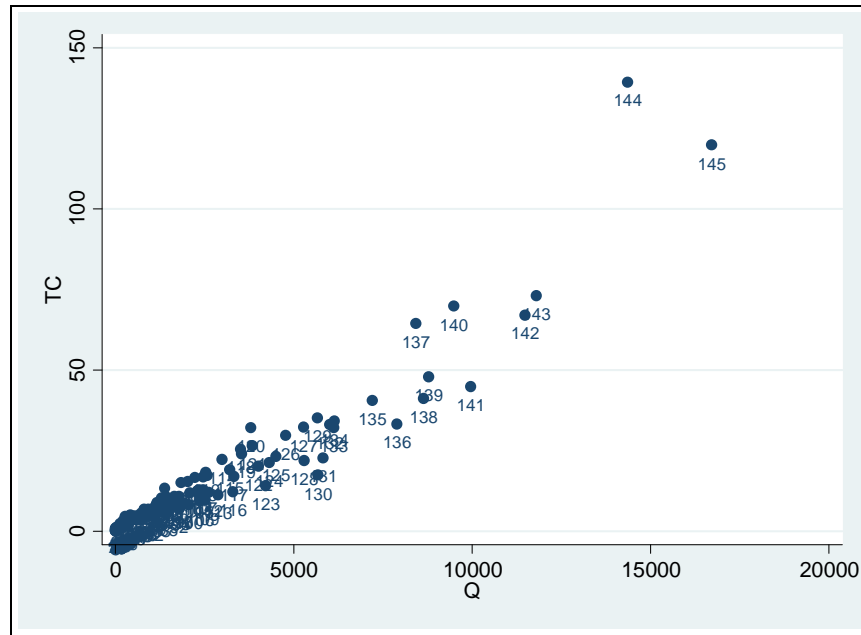
为此，首先定义一个新变量“ $n$ ”来表示第  $n$  个观测值。

```
. gen n=_n
```

其中，“ $\_n$ ”即表示第  $n$  个观测值。输入命令：

```
. scatter tc q,mlabel(n) mlabpos(6)
```

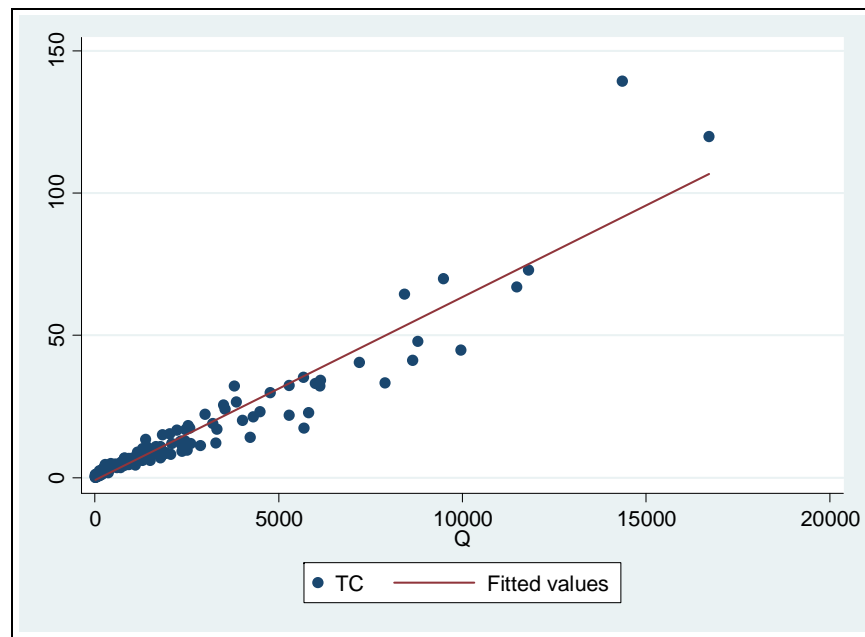
选择项“`mlabel(n)`”表示以变量“ $n$ ”作为“mark label”(标签)；选择项“`mlabpos(6)`”(mark label position)表示将此标签放在散点正下方(6点钟的位置)，默认位置为散点的右边(3点钟)。



如想在散点图上同时画出回归直线，使用命令：

```
. twoway (scatter tc q)(lfit tc q)
```

其中，“lfit”表示“linear fit”（线性拟合）。



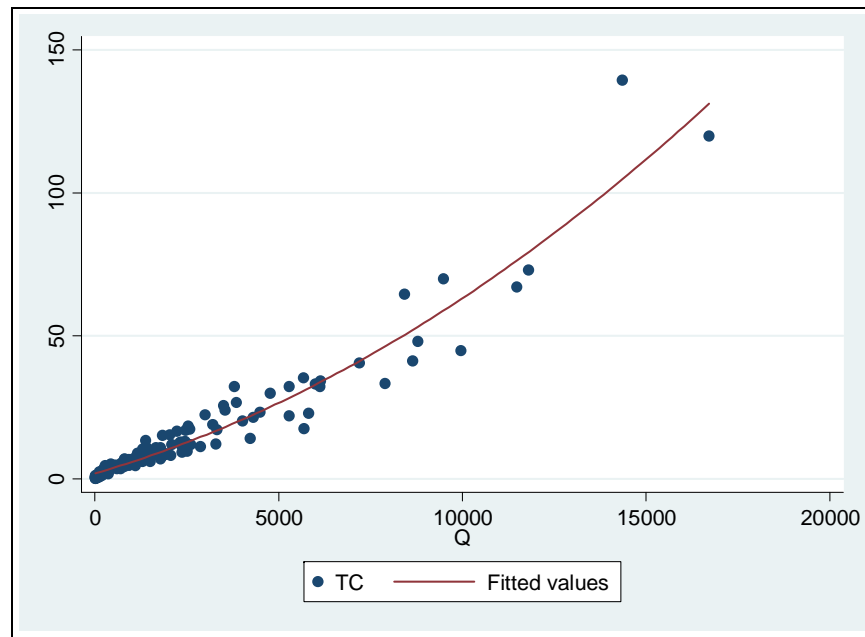
将此散点图存为文件名为“scatter1”的图像文件，以便调用。

```
. graph save scatter1  
(file scatter1.gph saved)
```

如想在散点图上同时画出二次回归曲线，使用命令：

```
. twoway (scatter tc q)(qfit tc q)
```

其中，“qfit”表示“quadratic fit”（二次拟合）。



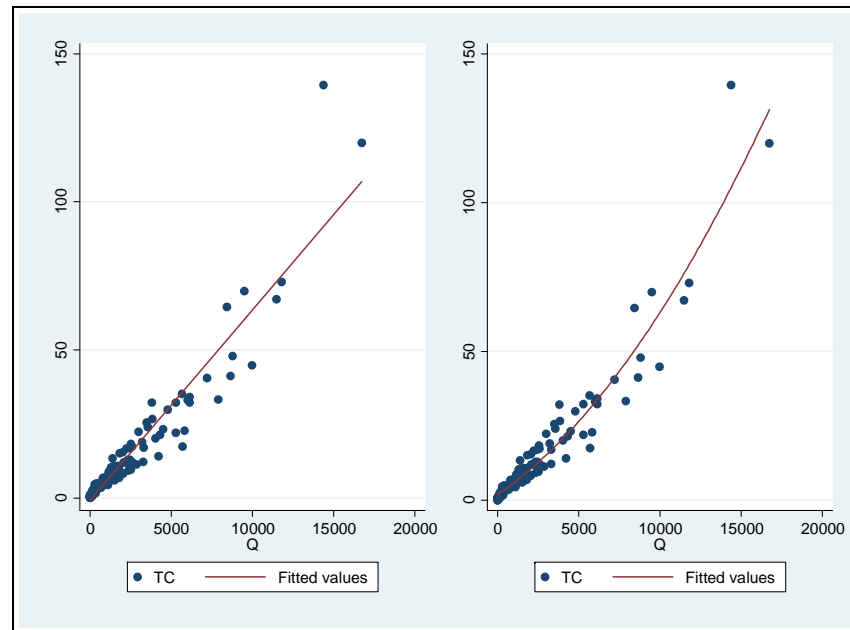
将此散点图存为文件名为“scatter2”的图像文件。

```
. graph save scatter2
```

```
(file scatter2.gph saved)
```

将上述两个图并列排放在一张图上。

```
. graph combine scatter1.gph scatter2.gph
```



更多作图方法，参见菜单“**Graphics**”。对于任何命令，只要输入“help command”（比如，help histogram），即可看到详细说明。

## 7. 生成新变量

Nerlove (1963)假设企业  $i$  的生产函数为 Cobb-Douglas 函数:

$$Q_i = A_i L_i^{\alpha_1} K_i^{\alpha_2} F_i^{\alpha_3}$$

$A, L, K, F$  分别为生产率、劳动力、资本与燃料。记  $r \equiv \alpha_1 + \alpha_2 + \alpha_3$  为规模效应(degree of returns to scale)。  $r = 1$ , 规模报酬不变;  $r > 1$ , 规模报酬递增;  $r < 1$ , 规模报酬递减。

假设企业追求成本最小化, 则成本函数为 Cobb-Douglas 函数:

$$TC_i = \delta_i Q_i^{1/r} (P_L)_i^{\alpha_1/r} (P_K)_i^{\alpha_2/r} (P_F)_i^{\alpha_3/r}$$

其中,  $\delta_i$  是  $A_i, \alpha_1, \alpha_2, \alpha_3$  的函数。取对数后得到,

$$\ln TC_i = \beta_1 + \frac{1}{r} \ln Q_i + \frac{\alpha_1}{r} \ln P_{L,i} + \frac{\alpha_2}{r} \ln P_{K,i} + \frac{\alpha_3}{r} \ln P_{F,i} + \varepsilon_i$$

在 Stata 中取对数，使用命令 `generate`。

```
. g lntc=log(tc)
. g lnq=log(q)
. g lnpl=log(pl)
. g lnpf=log(pf)
. g lnpk=log(pk)
```

如需要  $q$  的非线性平方项，使用命令

```
. g q2=q^2
```

如要生成 `lnpl` 与 `lnpk` 的互动项(interaction term)，使用命令

```
. g lnplpk=lnpl*lnpk
```



定义“ $q \geq 10000$ ”为大企业，并使用“虚拟变量”(dummy variable, 哑变量)large 来表示，

$$\text{large} \equiv \begin{cases} 1, & \text{如果 } q \geq 10000 \\ 0, & \text{其他} \end{cases}$$

可使用命令

```
. g larg=(q>=10000)
```

括弧“( )”表示对括弧中的表达式“ $q \geq 10000$ ”进行逻辑评估：如果为真，则取值为 1；如果为假，则取值为 0。

在上面命令中，不慎把 large 打成 larg 了。将变量重新命名：

```
. rename larg large
```

变量 larg 被重新命名为 large (也可使用变量管理器重新命名)。

假设想改变大企业的定义为“ $q \geq 6000$ ”，仍用 large 作为变量名。

方法一，先去掉现有变量 large，然后再定义一次：

```
. drop large  
. g large=(q>=6000)
```

方法二，更简洁的命令：

```
. replace large=(q>=6000)
```

将原变量( $q \geq 10000$ )直接替换为新变量( $q \geq 6000$ )。

某些变量名可能很长，一一输入变量名较费事。

方法一，直接在左下角的变量窗口单击需要的变量，该变量名就会显现在命令窗口。

方法二，如有以下变量 `lnq1`, `lnq2`, ..., `lnq30`，而只想使用其中的前 15 个变量，可用 `lnq1—lnq15` 来简略地表示这 15 个变量。

方法三，用 “\*” 号来节省变量名的书写。假设想将内存中所有以 “ln” 开头的变量都去掉，可输入命令

```
. drop ln*
```

这将去掉内存中的 `lntc`, `lnq`, `lnpl`, `lnpf`, `lnpk` 变量。

如果你后悔删除，Stata 并没有类似 Word 的 “undo” 命令，无法撤销此命令。

唯一的弥补方法是，重新使用命令 `generate`，再去生成这些变量。

## 8. Stata 的计算器功能

Stata 可作为计算器使用，命令格式 “display expression”。

计算 $\ln 2$ ：

```
. display log(2)  
.69314718
```

计算标准正态变量小于 1.96 的概率：

```
. di normal(1.96)  
.9750021
```

“normal” 表示标准正态的累积分布函数。常见概率分布的累积分布函数、密度函数等，参见 “help density function”。

## 9. 线性回归分析

使用 OLS 估计上述方程：

. regress lntc lnq lnpl lnpg lnpg

Source	SS	df	MS	Number of obs = 145		
Model	269.524728	4	67.3811819	F( 4, 140) =	437.90	
Residual	21.5420958	140	.153872113	Prob > F =	0.0000	
Total	291.066823	144	2.02129738	R-squared =	0.9260	
				Adj R-squared =	0.9239	
				Root MSE =	.39227	
lntc	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
lnq	.7209135	.0174337	41.35	0.000	.6864462	.7553808
lnpl	.4559645	.299802	1.52	0.131	-.1367602	1.048689
lnpg	-.2151476	.3398295	-0.63	0.528	-.8870089	.4567136
lnpg	.4258137	.1003218	4.24	0.000	.2274721	.6241554
_cons	-3.566513	1.779383	-2.00	0.047	-7.084448	-.0485779

“\_cons”表示常数项，“R-squared”显示 $R^2=0.9260$ ，“Adj R-squared”显示 $\bar{R}^2=0.9239$ 。

检验整个方程显著性的  $F$  统计量之  $p$  值(Prob > F)为 0.0000, 显示这个回归方程是高度显著的。

但  $\ln p_l$  与  $\ln p_k$  这两个变量均不显著, 其  $p$  值( $P > |t|$ )分别为 0.131 与 0.528。

变量  $\ln p_k$  的系数(Coef.)符号为负, 与经济理论的预测相反。Nerlove(1963)认为, 这是由于“资本使用成本”的数据不可靠。

表上方的回归结果显示, 残差平方和  $\sum_{i=1}^n e_i^2 = 21.542$ , 方程的标准误差(Root MSE)为  $s = 0.392$ 。

如果要显示估计系数的协方差矩阵，输入命令

. vce

Covariance matrix of coefficients of regress model					
e(V)	lnq	lnpl	lnpk	lnpf	_cons
lnq	.00030393				
lnpl	-.00035938	.08988127			
lnpk	.00034967	.02497537	.11548412		
lnpf	.00030089	-.01124831	-.00669535	.01006447	
_cons	-.00451909	-.15095534	-.59317676	.00784373	3.1662023

其中，“vce”表示“variance covariance matrix estimated”。

如果不要常数项，可以加上选择项 “noconstant”。

```
. reg lntc lnq lnpl lnpg lnpg,noc
```

如果只对“大企业”这个子样本进行回归，可输入命令

```
. reg lntc lnq lnpl lnpg lnpg if q>=6000
```

或者使用虚拟变量 large:

```
. reg lntc lnq lnpl lnpg lnpg if large
```

即只对“large=1”的子样本进行回归。



如想对“小企业”(除了“大企业”以外的所有企业)进行回归:

```
. reg lntc lnq lnpl lnpg lnpg if large==0
```

或者输入命令

```
. reg lntc lnq lnpl lnpg lnpg if ~large
```

其中,“~”表示逻辑的“否”(not)运算。

计算被解释变量的拟合值( $\hat{y}$ ), 并将其记为 lntchat:

```
. predict lntchat
```

计算“残差”(residual), 并将其记为 e1:

```
. predict e1, residual
```

选择项“residual”表示预测残差。如果没有选择项, “默认值”(default)计算拟合值 $\hat{y}$ 。

由于 lnq 的系数为 $1/r$ , 即规模报酬的倒数, 估计规模报酬为

```
. display 1/_b[lnq]  
1.387129
```

其中, “\_b[lnq]”表示“lnq”的 OLS 系数估计值。

由于  $\hat{r} = 1.387129 > 1$ ，故可能存在规模报酬递增。

检验规模报酬不变的原假设 “ $H_0 : r = 1$ ”：

. test lnq=1

此命令检验的原假设为，变量 lnq 的系数等于 1。

```
( 1)  lnq = 1  
  
      F(   1,   140) =   256.27  
      Prob > F =     0.0000
```

以很小的  $p$  值拒绝原假设，故认为存在规模报酬递增。

方程(4.3)显示, 变量  $\ln p_l$ ,  $\ln p_k$  与  $\ln p_f$  的系数之和等于 1。

```
. test (lnq=1)(lnpl+lnpk+lnpf=1)
```

```
( 1)  lnq = 1
( 2)  lnpl + lnpg + lnpg = 1

      F( 2, 140) = 128.15
      Prob > F = 0.0000
```

$p$  值 = 0.0000, 强烈拒绝此联合假设。

由于  $\ln p_l$  与  $\ln p_k$  均不显著，对二者的显著性进行联合检验：

```
. test lnpl lnpk
```

```
( 1)  lnpl = 0
( 2)  lnpk = 0

      F( 2, 140) =    1.69
      Prob > F =    0.1874
```

$p$  值很大(0.19)，可以接受二者的系数皆为 0 的联合假设。

Stata 也可检验“非线性假设”(参见第 5 章附录)。

例：检验变量 `lnpl` 的系数是 `lnq` 的系数的平方：

```
. testnl _b[lnpl]=_b[lnq]^2
```

```
(1)  _b[lnpl] = _b[lnq]^2

             F(1, 140) =          0.04
             Prob > F =          0.8334
```

由于  $p$  值很大(0.8334)，无法拒绝这个原假设。

## 10. 约束回归

经济理论要求  $\ln p_l$ ,  $\ln p_k$  与  $\ln p_f$  的系数之和为 1, 在此约束下重新估计原方程。

首先定义“约束条件 1”:

```
. constraint def 1  $\ln p_l + \ln p_k + \ln p_f = 1$ 
```

进行有约束的 OLS 估计:

```
. cnsreg lntc lnq  $\ln p_l$   $\ln p_k$   $\ln p_f$ , c(1)
```

其中, “cnsreg” 表示 “constrained regression”。

Constrained linear regression				Number of obs	=	145
				Root MSE	=	0.3915
( 1)   lnpl + lnpg + lnpg = 1						
lntc	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
lnq	.7213365	.0173912	41.48	0.000	.6869553	.7557176
lnpl	.6064693	.207239	2.93	0.004	.196772	1.016167
lnpg	-.0208375	.1933394	-0.11	0.914	-.4030563	.3613813
lnpg	.4143682	.0987832	4.19	0.000	.2190805	.6096559
_cons	-4.636069	.8949922	-5.18	0.000	-6.405408	-2.866731

变量 lnpg 的系数估计值从无约束 OLS 的“-0.22”变为约束 OLS 的“-0.021”，相对合理，但仍为负数，也依然不显著( $p$  值为 0.914)。



如果希望加上约束条件 “ $H_0:r=1$ ”，可定义 “约束条件 2”：

```
. cons def 2 lnq=1
```

在同时满足约束条件 1，2 的情况下进行回归：

```
. cnsreg lntc lnq lnpl lnpg lnpgf, c(1-2)
```

Constrained linear regression				Number of obs	=	145
				Root MSE	=	0.6553
( 1) $\ln pl + \ln pk + \ln pf = 1$						
( 2) $\ln q = 1$						
lntc	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
lnq	1	(constrained)				
lnpl	.1558956	.3436328	0.45	0.651	-.5234015	.8351927
lnpk	.1443526	.3231175	0.45	0.656	-.4943898	.7830949
lnpf	.6997518	.1626168	4.30	0.000	.3782892	1.021214
_cons	-7.926918	1.45791	-5.44	0.000	-10.80893	-5.044905

加上两个约束条件后，变量  $\ln pk$  的系数估计值符号终于变为正 (0.1443526)，但依然不显著( $p$  值为 0.656)；

另一方面，变量  $\ln pl$  变得不显著性( $p$  值上升为 0.651)。

## 11. Stata 的日志

如果希望在每次使用 Stata 时，储存其运行结果，可点击菜单“File”→“Log”→“Begin”，然后输入日志(log)的文件名，并存储在指定的位置。

在 Stata 中的所有操作及结果，都将记录日志中，直至选择退出。

如要暂时关闭日志(不再记录输出结果)，输入命令“log off”。

如果要恢复使用日志，可输入命令“log on”。

如果要彻底退出日志，只要输入命令“log close”即可。

如要查看日志文件的内容，点击存储位置上的日志文件图标。

## 12. Stata 命令运行结果的存储与调用

Stata 命令可以分为两种，即 e-类命令(e-class commands)与 r-类命令(r-class commands)。

e-类命令为“估计命令”(estimation commands)，比如“regress”；

所有其他命令为 r-类命令，比如，“summarize”。

r-类命令的运行结果存储在“r()”，可通过“return list”显示

```
. summarize q
```

Variable	Obs	Mean	Std. Dev.	Min	Max
q	145	2133.083	2931.942	2	16719

```
. return list
```

```
scalars:
      r(N) = 145
    r(sum_w) = 145
    r(mean) = 2133.08275862069
    r(Var) = 8596284.659770114
    r(sd) = 2931.942131040467
    r(min) = 2
    r(max) = 16719
    r(sum) = 309297
```

上表列出了在运行命令“summarize q”之后，Stata 所存储的结果，包括未显示的“r(Var)”(方差)、“r(sum)”(求和)等。

可调用这些结果作进一步的计算。比如，计算“变异系数” (coefficient of variation, 即标准差除以平均值):

```
. display r(sd)/r(mean)  
1.3745093
```

写得更漂亮些，可用命令:

```
. display "The coefficient of variation is "  
r(sd)/r(mean)  
The coefficient of variation is 1.3745093
```

e-类命令的运行结果存储在“e()”，输入“ereturn list”来显示

```
. reg tc q
```

Source	SS	df	MS	Number of obs = 145		
Model	51190.3707	1	51190.3707	F( 1, 143) = 1399.00		
Residual	5232.46776	143	36.5906836	Prob > F = 0.0000		
Total	56422.8385	144	391.825267	R-squared = 0.9073		
				Adj R-squared = 0.9066		
				Root MSE = 6.049		
tc	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
q	.0064307	.0001719	37.40	0.000	.0060908	.0067705
_cons	-.741095	.6219699	-1.19	0.235	-1.970538	.4883481

```
. ereturn list
```

```

scalars:
    e(N) = 145
    e(df_m) = 1
    e(df_r) = 143
    e(F) = 1399.000119377508
    e(r2) = .9072633016022542
    e(rmse) = 6.049023360142635
    e(mss) = 51190.37074066489
    e(rss) = 5232.467756451834
    e(r2_a) = .9066147932218505
    e(ll) = -465.7241671687837
    e(ll_0) = -638.1285147616334
    e(rank) = 2

macros:
    e(cmdline) : "regress tc q"
    e(title) : "Linear regression"
    e(marginsok) : "XB default"
    e(vce) : "ols"
    e(depvar) : "tc"
    e(cmd) : "regress"
    e(properties) : "b V"
    e(predict) : "regres_p"
    e(model) : "ols"
    e(estat_cmd) : "regress_estat"

matrices:
    e(b) : 1 x 2
    e(V) : 2 x 2

functions:
    e(sample)

```



上表列出了运行命令 `reg` 后 Stata 存储的结果，包括：

标量 (scalars)

宏 (macros)

矩阵 (matrices，即系数矩阵  $e(b)$  与协方差矩阵  $e(V)$ )

函数 (functions)。

## 4.4 Stata 命令库的更新

由于 Stata 版本不同(即使同为 Stata 12), 如果你发现本书中极少数命令无法运行, 可在命令窗口输入,

```
. update all
```

这将更新你的 Stata 命令库(Stata“ado”文件与其他可执行文件)。

Stata 用户还写了大量的外部命令或非官方命令(user-written software), 可直接下载到 Stata 中使用。

最流行的 Stata 非官方命令下载平台为“统计软件成分”(Statistical Software Components, SSC), 由 Boston College 维护, 网址为 <http://ideas.repec.org/s/boc/bocode.html>。

相关命令:

- . `ssc new` (罗列 SSC 的最新非官方 Stata 命令及简介)
- . `ssc hot` (罗列 SSC 提供的最流行非官方 Stata 命令)
- . `ssc install newcommand` (安装 SSC 非官方命令“newcommand”)
- . `help ssc` (有关 SSC 的帮助信息)

如使用“`ssc install newcommand`”下载非官方案序, 所有下载与安装过程将自动完成(包括新命令的帮助文件)。

如果要使用某种估计方法，但不知道它是否存在，可搜索

`. search keyword` (搜索帮助文件、FAQs、例子、*Stata Journal* (SJ), *Stata Technical Bulletin* (STB)等)

`. findit keyword` (搜索以上内容，以及 Stata 的网络资源)

命令 `findit` 的搜索范围比命令 `search` 更广些。

“`findit`” 等价于 “`search,all`”。

命令 `search` 的搜索结果较少，直接在 Stata 结果窗口显示

命令 `findit` 的搜索结果较多，将打开另一页面显示。

非官方命令的安装：

发现非官方命令后，如果不来自 `SSC`，一般需自行安装。

需要将所有相关文件下载到指定的 `Stata` 文件夹中(通常是 `ado\plus\`)。

如果不清楚应把文件复制到哪个文件夹，输入以下命令，以显示 Stata 的系统路径(system directories):

. sysdir

你会看到类似于以下的结果(取决于 Stata 的安装位置),

```
STATA:  D:\Stata12\  
UPDATES: D:\Stata12\ado\updates\  
  BASE:  D:\Stata12\ado\base\  
  SITE:  D:\Stata12\ado\site\  
  PLUS:  c:\ado\plus\  
PERSONAL: c:\ado\personal\  
OLDPLACE: c:\ado\
```

将下载的新命令文件复制到 PLUS 所指示的那个文件夹即可(此处为 “c:\ado\plus\” )。

## 4.5 进一步学习 **Stata** 的资源

更多 Stata 知识，将在本书以后章节中逐步介绍。

Stata 英文参考书：Baum (2006)，Cameron and Trivedi (2009)，以及 Stata 出版社(Stata Press)出版的系列书籍。

加州大学洛杉矶分校 (UCLA) 网站 (<http://www.ats.ucla.edu/stat/stata/>)有大量 Stata 的资源及实例(搜索“Stata UCLA”即可找到此网站)。

中文参考书包括陈传波《Stata 十八讲》，胡咏梅(2010)，兰草(2012)，劳伦斯·汉密尔顿(2008)，李春涛、张璇(2009)，王群勇(2007, 2008)，王天夫、李博柏(2008)，杨菊华(2012)，张鹏伟、李嫣怡(2011)等。

Stata 本身的“帮助”(Help)菜单包含了详细的信息，比如，“`help reg`”。

更进一步的学习，可查看 Stata 手册(Stata manuals)。

在 Stata 11 中，每个命令的帮助页面(比如“`help reg`”)底部均有相应的 Stata 手册链接。