

## 第八章 Stata

### §8.1 应用概要

#### §8.1.1 简介

Stata 是一个非常实用的统计软件包，具有数据显示与管理、统计分析和统计作图等功能，与电子报表程序、数据库管理程序、统计软件包、图形软件包和程序语言有着共同点。它的设计注重实用，系统小从而不但使用方便，而且功能强大。从3.0版以后功能大大增强。从4.0开始，有了Windows产品，可以全屏幕浏览数据，仅仅使用鼠标器便从review窗口和变量窗口挑选用过的命令和变量，十分方便，输出结果可以将命令醒目地标示出来。另外，其宏定义ADO对用户是透明的，特别适于研究者实现新的分析方法。5.0中增加了长期数据分析和复杂调查分析等内容。

Stata 可用于各种计算机，如IBM PC、PS/2, 80386、68000、VAX/VMS 和UNIX 工作站，但各种机型的文件可以自动转换。

#### §8.1.2 系统运行

##### (一)、进入和退出

建议DOS 的CONFIG.SYS 文件中增加语句DEVICE=[d:][path]ANSI.SYS，这样在软件使用时，将显示以彩色。在DOS 提示符下，打入

```
D:\Stata>stata
```

出现系统提示符(.)。

Windows版需要运行wstata。使用命令行参数/kxxx可以指示所占用内存的大小。stata 支持数学协处理器和EMS。

Stata 只识别以小写字母打入的命令。

退出仅需打入exit 命令(放弃改动过的数据时用exit,clear 命令)。

##### (二)、启用帮助和菜单

在提示符下打入help (F1+Enter) 命令，软件显示Stata 的全部命令，这一点与FoxBASE+有些类似。在新版本Stata中，使用help contents给出详细的命令分类，使用lookup可以查找命令所关联的产品。

编辑键：Ctrl-字母键的组合：W(End)、E(Insert)、R(estore)、U(Esc)、J(Enter)、M(Enter)、D(elete)、G(Begin of Word)、O(Being of line)、P(End of line)。在SUN 计算机上尚有R1-R15键。其它的如↑、↓、←、→移动，Ins插入、Del删除。行首(Home)、行尾(End)、上一行(PgUp)、下一行(PgDn)、跳格(Tab)、寻找(Ctrl-Home 加行号)。可用#review [n] 显示已入过的命令。Windows 版功能键略有不同。

在安装了stata/mnu时，可用F10 进入菜单系统。菜单项以光标键和跳格键结合回车键进行选择。在Windows版本中使用edit和browse编辑和浏览数据工作表。

##### (三)、语言特色

Stata 的数据存贮类型有五种，即整型、单精度或双精度浮点数。5、-5、5.2、.5、5.2E+2、5.2D+2均是合法的表示。有以下的表格供参考：

字符串存贮类型有str2,str4,...,str80，如"Example"、"2.1"、""。

Stata 的名称是1-8 个字母、数字、下划线，对大小写是敏感的。Stata保留使用以下变量：

`_all`(所有变量)、`float`(指示浮点类型)、`_rc`(返回码)、`_skip`(跳记录)、`_b`(回归系数)、`if`(条件)、`_n`(当前记录)、`using`(使用文件名)、`_coef`(系数)、`in`(观察范围)、`_N`(总观察)、`_weight`(体重)、`_cons`(常数)、`int`(整型指示)、`_pi`( $\pi$ )、`with`(与)、`double`(双精度)、`long`(长整数)、`_pred`(预测值)。

Stata 的原始数据是一个矩形的表格，不同的变量以变量名表示观测值由1至N。Stata 同时使用交叉乘积类型的资料。设原始数据矩阵为X，第一列由系统变量`_cons`组成，其每个元为0， $X'X$ 称为XP形式，可由命令`convert`形成。`set contents xp`可以把第一个量改名为`_cons`。

Stata 的使用以下的文件指示：

```
.dct ASCII 数据字典
.do 运行文件
.dta Stata 格式的数据文件
.gph 图形文件
.log 运行记录文件
.raw ASCII 格式的数据集
.XP Stata 格式的交叉乘积数据集
```

文件操作具有基本格式：`<文件操作关键字> [变量列表] using <文件名>`，如：

```
infile x1 x2 str10 (x3 x4) y1-y10 using myin
outfile using myout
graph using mygph
log using mylog
```

第一句和第二句分别读入文件`myin`和写出文件`myout`，第三句调入并显示图形文件`mygph`，第四句用`mylog`文件存放结果，打开以后，以LOG ON和LOG OFF作记录运行信息的切换开关。

除少数例外，Stata 使用的语句格式为：

`[by varlist] 命令[varlist] [=exp] [if exp] [in range] [,options]`。其中`varlist`为变量列表。`exp`为施加的权，`exp`是一个容纳条件的表达式，`range`表示范围，`options`是有关的选项，一般来说`options`与`nooptions`对应。某些选项含有参量应该放在括号内。

Stata 把开头为星号(\*)的行视为注释，在DO文件中可用`/* */`表示。

Stata 的表达式与其它软件的法并没有区别，与C语言类似。

数学函数：`abs(x)`、`acos(x)`、`asin(x)`、`atan(x)`、`comb(x)`、`cos(x)`、`exp(x)`、`ln(x)`、`lnfact(x)`、`lngamma(x)`、`log(x)`、`log10(x)`、  
统计函数：

```
Binomial(n,k, $\pi$ ) 在n次观察中k次以上的概率
binorm(h,k, $\rho$ ) 相关系数为  $\rho$  的两维正态累积分布
chiprob(df,x)、invchi(df,p) 自由度为x的累积卡方。
fprob(df1,df2,f)、invfprob(df1,df2,p) F-分布。
gammamp(a,x)、invgammamp(a,p) 不完全伽马分布
ibeta(a,b,x) 不完全贝塔分布
invbinomial(n,k,p) 逆二项分布
```

invnchi(df,l,p)、nchi(df,l,x)、npnchi(df,x,p) 非中心卡方分布。

invnorm(p)、normd(z)、normprob(z) 正态分布。

invt(df,p)、tprob(df,t) t分布。

uniform()均匀伪随机数

日期函数: date(s1,s2)、day(e)、dow(e)、mdy(m,d,y)、month(e)、year(e)。

字符串函数: index(s1,s2)、length(s)、lower(s)、ltrim(x)、real(s)、rtrim(x)、string(n)、substr(s,n1,n2)、trim(x)、upper(s)

其他函数: autocode(x,ng,xmin,xmax)、cond(x,a,b)、float(x)、group(x)、int(x)、max(x1,...,xn)、min(x1,...,xn)、sum(x)、, xn)、round(x,y)、sign(x)、sum(x)。

Stata 编程首先要了解参数的宏传递的, 所以可用命令'help macros' 获得一些信息, 如程序名后第一个参数是%\_1, 第二个是%\_2, 等等。设数据集中的缺失值记为-9, 要把它们改为Stata 的缺失值, 相当于命令:

```
replace varname=. if varname==-9
```

但若有十个这样的变量, 打很多这样的命令是很烦琐的。借助程序

```
program define fix
  replace %_1=. if %_1==9
end
```

就产生了名为fix 的程序名, 此后仅仅使用'fix var1', 就相当于执行: 'replace var1=. if var1==9', 其余的情形依次类推。操作如下:

```
. program define fix
1. replace %_1=. if %_1==9
2. end
```

设原始数据变量为z, 则命令fix z 启用fix的操作。

#### (四)、运行实例

Stata 提供了一套较完整的教学程序, 如: intro.tut,graphics.tut,tables .tut,anova.tut,regress.tut,probit.tut survive.tut, statkit.tut, graphkit. tut, datakit.tut, qckit.tut, ourdata.tut, yourdata.tut.使用do或tutorial 命令。如intro.tut的分析:

```
. use c:\stata\auto
. describe
. drop mpg hdroom runk length turn displ gratio
. list in 1
. summarize rep78
. tabulate rep78, plot
. list if rep78==1
. list if rep78==5
. tabulate rep78 foreign
. tabulate rep78 foreign, column
. tabulate rep78 foreign, column nofreq chi2
. tabulate foreign, summarize(rep78)
```

```

. graph weight price
. plot weight price
. correlate weight price
. regress weight price
. predict resid, residual
by foreign: summarize weight price resid
. generate weightd = weight if ~foreign
. generate weightf = weight if foreign
. graph weightd weightf price
. plot weightd weightf price
. graph weightd weightf price
. graph price weight, by(foreign) total
. plot weightd weightf price
. regress weight price foreign

```

变量rep78 是消费者的报告维修排名，用summarize 来观察，其次是最坏的车型和最好的车型。结果好的车型似以外国车为主，平均来说是这样的吗？变量'foreign' 是一个示性量，若车为国产，其值为0 否则为1，借助它对两种类型行制表。使用绝对数和百分比，并计算国产车与外国车之间的卡方统计量，而且比较两者之间的平均维修记录。另外看车重与车的价格，关系如何？一种方式是通过绘图来观察。图形显示，随着价格的上升，车重也增加了。没有图形设备时，使用plot 图示效果也一样。可以算得车重与价格的相关系数，预计会是一个正值。进一步，有车重与价格的回归分析。回归的残差用一个名为'resid' 的量来表示。回归分析之后，Stata 存贮了这些变量以及系数，predict 则使用了这些信息，结合原始数据来计算predictions, residuals, 以及influence statistics。这样可以观察，国产与进口车维修记录有所不同，在车重/价格关系上也是这样的吗？图形显示，同一价格下的国产车几乎均比进口车要重。

## §8.2 统计分析

### §8.2.1 统计制表

命令为: tabulate

58 个病人的材料，它们患有三种病，服用四种药，记录他们血压的变化。原始数据为：

	Disease 1	Disease 2	Disease 3
Drug 1	42,44,36 13,19,22	33,26,33 21	31,-3,25 25,24
Drug 2	28,23,34 42,13	34,33,31 36	3,26,28 32,4,16
Drug 3	1,29,19	11,9,7 1,-6	21,1,9 3
Drug 4	24,9,22 -2,15	27,12,12 -5,16,15	22,7,25 5,12

```
. use systolic, clear
```

```
. tabulate drug
. tabulate drug disease
```

使用命令tabulate 产生列表。在两维表中，Stata 可以有許多选择项，如nofreq (不打印频数)、column (报告列百分比)、row (报告行百分比) 及cell (格子百分比)。下表显示了disease 与drug 交叉表有关统计量，命令、选项和变量名都略写了。

```
. tab dis dr, col row cell
. tab disease drug, chi2
```

Patient's  Drug Used						
Disease	1	2	3	4	Total	
1	6	5	3	5	19	
2	4	4	5	6	19	
3	5	6	4	5	20	
Total	15	15	12	16	58	

```
chi2(6) = 1.4048 Prob>chi2 = 0.966
```

对于多维表格也类似，如关于三个量var1、var2 及var3, 使用：

```
. sort var1
. by var1: tabulate var2 var3
```

每种药物与收缩压的关系列表。

```
. tabulate drug, summarize(systolic)
      | Summary of Increment in Systolic B.P.
Drug Used|      Mean   Std. Dev.      Freq.
-----+-----
      1 |      26.07    11.68         15
      2 |      25.53    11.62         15
      3 |       8.75    10.02         12
      4 |      13.50     9.32         16
-----+-----
      Total |      18.88    12.80         58
. tab disease drug, sum(systolic)
. tab disease drug, sum(systolic) mean
. sort var1
. by var1: tab var2 var3, sum(contvar)
```

产生分组统计量，结果包括了均值、标准差及收缩压增加频数或者仅仅显示其中某个统计量(mean, standard, freq)，多维也类似。

## §8.2.2 方差与协方差分析

进行方差分析使用命令 `oneway`、`anova` 和 `test`。现仍然采用制表部分所提供的58个病人的材料。`oneway` 命令估计 one-way ANOVA 模型并能做两两比较，其语法为：`oneway` 反应变量控制变量，此处

```
. oneway systolic drug
. oneway systolic drug, tabulate
. oneway systolic drug, bonferroni
. test drug
. test drug drug*disease
. test drug, error(drug*disease)
. anova systolic drug disease
. regress
. anova systolic drug disease
. anova systolic disease drug, sequential
. anova systolic drug disease drug*disease
. test _coef[ drug[2] ] = _coef[ drug[3] ]
. test _coef[disease[2]] + 3*_coef[disease[3]] = 6 + _coef[disease[3]]
. quietly anova systolic drug disease drug*disease
. test, symbolic
. test drug, symbolic
. use sysage, clear
. summarize age
. anova systolic drug disease age disease*age, continuous(age)
```

程序结果如下：

Analysis of Variance					
Source	SS	df	MS	F	Prob > F
Between groups	3133.23851	3	1044.41284	9.09	0.0001
Within groups	6206.91667	54	114.942901		
Total	9340.15517	57	163.862371		

Bartlett's test for equal variances:  $\chi^2(3) = 1.0063$  Prob> $\chi^2 = 0.800$

Summary of Increment in Systolic B.P.			
Drug Used	Mean	Std. Dev.	Freq.
1	26.07	11.68	15
2	25.53	11.62	15
3	8.75	10.02	12

4	13.50	9.32	16
-----+			
Total	18.88	12.80	58

Bartlett's test for equal variances:  $\chi^2(3) = 1.0063$  Prob> $\chi^2 = 0.800$   
 Comparison of Increment in Systolic B.P. by Drug Used  
 (Bonferroni)

Row Mean-			
Col Mean	1	2	3
----- -----			
2	-0.53		
	1.000		
3	-17.32	-16.78	
	0.001	0.001	
4	-12.57	-12.03	4.75
	0.012	0.017	1.000

Number of obs = 58      R-square = 0.4560  
 Root MSE = 10.5096      Adj R-square = 0.3259

Source	Partial SS	df	MS	F	Prob > F
-----+					
Model	4259.33851	11	387.212591	3.51	0.0013
drug	2997.47186	3	999.157287	9.05	0.0001
disease	415.873046	2	207.936523	1.88	0.1637
drug*disease	707.266259	6	117.87771	1.07	0.3958
Residual	5080.81667	46	110.452536		
-----+					
Total	9340.15517	57	163.862371		

Source	Partial SS	df	MS	F	Prob > F
-----+					
drug	2997.47186	3	999.157287	9.05	0.0001
Residual	5080.81667	46	110.452536		
-----+					
Source	Partial SS	df	MS	F	Prob > F
-----+					
drug drug*disease	3770.69912	9	418.966569	3.79	0.0012
Residual	5080.81667	46	110.452536		
-----+					
Source	Partial SS	df	MS	F	Prob > F

```

-----+-----
                drug | 2997.47186    3  999.157287    8.48    0.0141
drug*disease | 707.266259    6  117.87771
-----+-----
Source |      SS      df      MS                Number of obs =      58
-----+-----                F( 5, 52) =      6.38
Model | 3552.07225    5  710.414449          Prob > F      = 0.0001
Residual | 5788.08293   52  111.309287          R-square      = 0.3803
-----+-----                Adj R-square  = 0.3207
Total | 9340.15517   57  163.862371          Root MSE     = 10.55

```

Response variable is systolic

```

-----+-----
Variable   Coefficient   Std. Error   t   Prob > |t|   Mean
-----+-----
_cons      9.878751     3.317476    2.978  0.004        1
drug
  1        12.46897     3.807342    3.275  0.002     .2586207
  2        12.36457     3.80698    3.248  0.002     .2586207
  3       -4.526825     4.033439   -1.122  0.267     .2068966
  4        (dropped)
disease
  1         6.434081     3.388837    1.899  0.063     .3275862
  2         4.294931     3.400646    1.263  0.212     .3275862
  3        (dropped)
-----+-----

```

( 1) drug[2] - drug[3] = 0.0, F(1, 52) = 16.89, Prob > F = 0.0001

( 1) disease[2] + 2.0 disease[3] = 6.0, F(1,52) = 0.25, Prob > F= 0.6182

```

                Number of obs =      58          R-square      = 0.6221
                Root MSE     = 8.48737          Adj R-square  = 0.5604

```

```

Source | Partial SS   df      MS      F      Prob > F
-----+-----
Model | 5810.41855   8  726.302318  10.08  0.0000
|
drug | 2791.94475   3  930.648251  12.92  0.0000
disease | 129.092158   2  64.5460789  0.90  0.4148
age | 1817.80067   1  1817.80067  25.23  0.0000
disease*age | 43.4069507   2  21.7034754  0.30  0.7412
|

```



```

Residual | 3529.73663    49  72.0354414
-----+-----
Total | 9340.15517    57  163.862371

```

算得的组间方差提示，接受不同药物的病人之间收缩压的改变是显著的。使用tabulate选项时，Stata报告控制变量每个水平上的均值，其余结果相同。本例用Bonferroni方法进行两两比较的方法有。首先指示一个两因素、没有交互的模型，反应变量是血压，控制变量是药物和疾病种类。检验中用的平方和是偏平方和。使用test语句检验药物的作用以及交互作用：除非特别指定，Stata使用的是用剩余均方作为误差项，否则要进行特指。方差分析之后，又能得到回归。test语句也可用于检验回归系数，此处检验第二、第三种药物的系数是否相同。Stata还可以给出用记号表示的函数和特定检验的可估计函数。在明确那些变量是连续的之后，可进行协方差分析。以上采用的协变量是age。

### §8.2.3 回归分析

使用语句：regress、test、predict和stepwise。现用人口普查的资料进行说明。运行的程序为：

```

. describe
. summarize
. regress drate medage medagesq pcturban
. test medage medagesq
. test medage=2*medagesq
. test 2*(medage-medagesq)-(medage-medagesq)/2=(medage-medagesq)/2+medagesq
. test pcturban=medage, accumulate
. correlate, _coef
. stepwise drate medage medagesq pcturban
. predict dhat
. summarize drate dhat
. predict influ, cooksd
. summarize influ, detail
. list state if influ>1
. regress drate medage medagesq pcturban if influ<1
. summarize pop
. regress drate medage medagesq pcturban =pop
. use hsng, clear
. keep hsngval faminc rent pcturban region
. describe
. regress rent hsngval pcturban (faminc reg2-reg4 pcturban)

```

其结果为：

```

Source |      SS      df      MS      Number of obs =      50
-----+-----
F( 3, 46) = 31.47

```

Model		.00005593	3	.000018643	Prob > F	=	0.0000
Residual		.000027249	46	5.9236e-07	R-square	=	0.6724
-----+-----							
Total		.000083179	49	1.6975e-06	Adj R-square	=	0.6510
					Root MSE	=	.00077

Variable		Coefficient	Std. Error	t	Prob >  t	Mean
-----+-----						
drate						.008436
-----+-----						
medage		.0004851	.001207	0.402	0.690	29.54
medagesq		2.37e-06	.0000206	0.115	0.909	875.422
pcturban		-.0000353	8.29e-06	-4.262	0.000	66.94913
_cons		-.005598	.0178979	-0.313	0.756	1
-----+-----						

( 1) medage = 0.0  
 ( 2) medagesq = 0.0  
 F( 2, 46) = 44.03, Prob > F = 0.0000  
 ( 1) medage - 2.0 medagesq = 0.0  
 F( 1, 46) = 0.15, Prob > F = 0.7021  
 ( 1) medage - 2.0 medagesq = 0.0  
 F( 1, 46) = 0.15, Prob > F = 0.7021  
 ( 1) medage - 2.0 medagesq = 0.0  
 ( 2) - medage + pcturban = 0.0  
 F( 2, 46) = 21.85, Prob > F = 0.0000

		medage	medagesq	pcturban	_cons
-----+-----					
medage		1.0000			
medagesq		-0.9985	1.0000		
pcturban		0.3235	-0.3352	1.0000	
_cons		-0.9984	0.9942	-0.3385	1.0000

Variable		Obs	Mean	Std. Dev.	Min	Max
-----+-----						
drate		50	.008436	.0013029	.0039915	.0106902
dhat		50	.008436	.0010684	.0045433	.0111047

Source		SS	df	MS	Number of obs	=	50
-----+-----							
Model		.000038537	3	.000012846	F( 3, 46)	=	40.83
Residual		.000014471	46	3.1458e-07	Prob > F	=	0.0000
					R-square	=	0.7270

-----+-----					Adj R-square = 0.7092	
Total		.000053008	49	1.0818e-06	Root MSE = .00056	
Variable		Coefficient	Std. Error	t	Prob >  t	Mean
-----+-----						
drate						.0087368
-----+-----						
medage		.0005844	.000994	0.588	0.559	30.11047
medagesq		-4.36e-07	.0000163	-0.027	0.979	909.3716
pcturban		-.0000285	6.53e-06	-4.368	0.000	73.66408
_cons		-.0063644	.0152071	-0.419	0.678	1
-----+-----						

Dropping: medagesq F= 0.0133

(stepwise)

Source		SS	df	MS	Number of obs = 50
-----+-----					
Model		.000055922	2	.000027961	F( 2, 47) = 48.22
Residual		.000027256	47	5.7993e-07	Prob > F = 0.0000
-----+-----					
Total		.000083179	49	1.6975e-06	R-square = 0.6723
-----+-----					
					Adj R-square = 0.6584
					Root MSE = .00076

Variable		Coefficient	Std. Error	t	Prob >  t	Mean
-----+-----						
drate						.008436
-----+-----						
medage		.0006238	.0000658	9.483	0.000	29.54
pcturban		-.000035	7.73e-06	-4.531	0.000	66.94913
_cons		-.0076466	.0019034	-4.017	0.000	1
-----+-----						

(2SLS)

Source		SS	df	MS	Number of obs = 50
-----+-----					
Model		36677.4033	2	18338.7017	F( 2, 47) = 42.66
Residual		24565.7167	47	522.674823	Prob > F = 0.0000
-----+-----					
Total		61243.12	49	1249.85959	R-square = 0.6448
-----+-----					
					Adj R-square = 0.6297
					Root MSE = 22.862

Variable		Coefficient	Std. Error	t	Prob >  t	Mean
-----+-----						
rent						234.76
-----+-----						

hsngval	.0022398	.0003388	6.612	0.000	48484
pcturban	.081516	.3081528	0.265	0.793	66.94913
_cons	120.7065	15.70688	7.685	0.000	1

-----+-----

本例用regress 命令进行线性回归。回归方程为：

$$\text{drate} = b_0 + b_1 \text{medage} + b_2 \text{medagesq} + b_3 \text{pcturban}$$

drate 是州死亡率，medage 是州人口龄中位数，medagesq 是年龄平方中位数，pcturban 是生活在城区的人口比例。回归结束后，使用regress 命令可以重显回归的结果。使用test 命令进行有关的检验，若指定accumulate 选项后，可以检验多个假设。估计量的协方差阵可以使用'correlate, \_coef'而得。逐步回归方法有向后(backward)、向前(forward)、逐步(stepwise)，进入和剔除的F-值由fenter(#) 和fstay(#) 指定，默认值分别为0.5 和0.1。

其它的量有：预测值、残差、标化残差、学生化残差、预测标准误、影响统计量(Cook's 距离、投影阵对角元和DF-Betas)。一个观察是高影响点，现把它找出来。除掉该点继续估计。regress 命令也能进行加权回归，现在数据中含变量pop 表示每处的人口数，最后一个是工具变量法和两阶段最小二乘解。语句格式为regress lhsvar rhsvar1 rhsvar2 ... (exogvar1 exogvar2 ... )，之后进行假设检验或计算预测值了。

数据hsng 记录了1980 人口普查数据分析(median housing values 及rents)，估计的模型是：

$$\text{hsngval} = a_0 + a_1 \text{faminc} + a_2 \text{reg2} + a_3 \text{reg3} + a_4 \text{reg4}$$

$$\text{rent} = b_0 + b_1 \text{hsngval} + b_2 \text{pcturban}$$

#### §8.2.4 logit/probit 分析

命令probit 和logit，仍然采用汽车数据示范。

```
. use auto, clear
. keep make foreign mpg weight
. describe
. inspect foreign
. logit foreign weight mpg
. logit, tabulate nocoef
. probit foreign weight mpg
. predict probhat
. summarize probhat
. list in 1/13
. predict zhat, index
. summarize probhat zhat
. correlate, _coef
. cor, _c cov
. test (weight-mpg)/2 - (mpg-weight) = mpg - (weight+mpg)*2
. test mpg=0, accumulate
. probit foreign weight mpg if mpg>18
. tabulate foreign repair
. probit foreign repair1 repair2
```

运行结果:

```
foreign: Car type
-----
                Number of Observations
                Total   Integers   Non-
                -----   -----   -----
| #             Negative   -         -         -
| #             Zero       52        52        -
| #             Positive   22        22        -
| #             -----   -----   -----
| # #           Total     74        74        -
| # #           Missing   -         -         -
+-----+-----+-----+
0             1             74
(2 unique values)
```

```
Logit Estimates                Number of obs =    74
                                chi2(2)         =   35.72
Log Likelihood =-27.175156     Prob > chi2    = 0.0000
```

Variable	Coefficient	Std. Error	t	Prob >  t	Mean
foreign					.2972973
weight	-.0039067	.0010116	-3.862	0.000	3019.459
mpg	-.1685869	.0919175	-1.834	0.071	21.2973
_cons	13.70837	4.518709	3.034	0.003	1

#### Comparison of Outcomes and Probabilities

Outcome	Pr < .5	Pr >= .5	Total
Failure	46	6	52
Success	9	13	22
Total	55	19	74

```
Probit Estimates                Number of obs =    74
                                chi2(2)         =   36.38
Log Likelihood =-26.844189     Prob > chi2    = 0.0000
```

Variable	Coefficient	Std. Error	t	Prob >  t	Mean
----------	-------------	------------	---	-----------	------

```

-----+-----
foreign |                                     .2972973
-----+-----
weight |   -.0023355   .0005661   -4.126   0.000   3019.459
  mpg |   -.1039503   .0515689   -2.016   0.048   21.2973
  _cons |    8.275464    2.554142    3.240   0.002     1
-----+-----

```

```

Variable |   Obs       Mean   Std. Dev.   Min       Max
-----+-----
probhat |    74   .294487   .3074146   9.52e-06   .9029781
  zhat |    74  -.9904795   1.376307  -4.275976   1.298709

```

```

      | weight   mpg   _cons
-----+-----
weight|  1.0000
  mpg |  0.7809  1.0000
  _cons| -0.9504 -0.9305  1.0000
      | weight   mpg   _cons
-----+-----
weight|  3.2e-07
  mpg | .000023  .002659
  _cons| -0.001374 -0.122554  6.52364

```

( 1) 3.5 weight - .5 mpg = 0.0, F(1, 71) = 3.26, Prob > F = 0.0753

( 1) 3.5 weight - .5 mpg = 0.0

( 2) mpg = 0.0

F( 2, 71) = 10.38

Prob > F = 0.0001

Probit Estimates

Number of obs = 47

chi2(2) = 23.10

Log Likelihood = -19.203993

Prob > chi2 = 0.0000

```

Variable | Coefficient   Std. Error   t   Prob > |t|   Mean
-----+-----
foreign |                                     .3617021
-----+-----
weight |   -.0023289   .0007581   -3.072   0.004   2642.766
  mpg |   -.0302598   .060018   -0.504   0.617   24.34043
  _cons |    6.14531    3.02718    2.030   0.048     1
-----+-----

```

```

Probit Estimates                                Number of obs =    48
                                                chi2(1)          =    9.53
Log Likelihood =-22.229138                    Prob > chi2     = 0.0020

```

Variable	Coefficient	Std. Error	t	Prob >  t	Mean
foreign					.25
repair2	-1.281552	.4297326	-2.982	0.005	.625
_cons	-3.75e-16	.295409	-0.000	1.000	1

logit 和probit 用于分析属性因变量的情形，这里用是否国产车或进口车作为分析变量，从inspect我们明确分析变量具体的取值。foreign 加了标号，所有取值存在标号里。取协变量为车重和里数做回归，计算开始时对数似然值为-45.03321，迭代5次后收敛于-27.175156，继续用logit 可以重显计算结果。probit 和logit 命令都可以产生一个原始分类与计算分类的比较表，这由tabulate选项完成。probit 计算迭代初值为亦为-45.03321，5次后的结果为-26.844189。probit 和logit 均由迭代技术求取非线性估计量，收敛性可由三个选项控制。iterate(#) 指最大迭代次数，tolerance(#) 指示系数容许性，ltolerance(#) 指示对数似然函数的容许性。predict 命令同样可以用于预测，这里的预测结果是概率值。predict 也能计算index 函数的预期值，对于probit 分析，预测概率是 $\text{probhat} = F(\text{zhat})$ ， $F()$  是累积正态分布。Stata 的probit 和logit 命令与其它命令的特性一样，模型分析之后，可以得到估计量基于信息矩阵的协方差阵，即信息阵的逆阵。根据这些信息进行检验，accumulate 选项允许检验的累积，因而是联合检验。其它的特征，如指示子集分析，对数似然函数迭代初值为-30.756389，5次收敛于-19.203993。特别重要的是，当模型中的因变量完全决定了结果的成功与否，那么许多程序会算出无穷大的系数，或者产出不过是四舍五入误差的解，Stata 能及时以现这种情形。汽车数据增加了一个名为repair 的量，取值为1, 2, 3, 表示差、平均和好。由于数据中维修记录最差的均是国产车，若要用维修记录来预计外国厂商，则 $\text{repair}=1$  的概率为0，即probit 或logit 的系数会无穷小。而Stata 的结果正确。迭代自-26.992087 始，4次收敛于-22.229138。

### §8.2.5 生存分析

使用命令kapmeier, gwood, mantel, wilcoxon, logrank, survcurv, loglogs, cox, coxvar, coxbase, boxhaz. cox 命令能对有关时间、截尾和时变协变量资料估计比例风险模型，现用一个发电机的资料，比较新旧类型的轴承，记录这些发电机在超负荷下运行直至损坏的小时数。

```

. use kva, clear
. describe
. cox failtime load bearings
. correlate, _coef
. correlate, _coef covariance

```

运行结果：

```

Cox regression                               Number of obs =    12
                                              chi2(2)          =   23.39
Log Likelihood = -8.577853                  Prob > chi2     = 0.0000

```

Variable	Coefficient	Std. Error	t	Prob >  t	Mean
failtime					74.66667
load	.4229578	.1433485	2.951	0.015	27.5
bearings	-2.754461	1.173115	-2.348	0.041	.5

```

          |      load bearings
-----+-----
load |   .020549
bearings | -.121008   1.3762

```

对数似然迭代初值为-20.274897, 经5次迭代收敛至-8.577853。load 与bearings 间的相关为-0.7196。不带参数执行cox 命令时, 也会重显回归的结果。

估计带有截尾观察的数据, 使用的是一个药物实验的数据。

```

. use cancer, clear
. describe
. tabulate died, summarize(studytim)
. tabulate drug, summarize(studytim), if died
. tabulate drug, summarize(died)
. quietly tabulate drug, gen(drug)
. cox studytim age drug2 drug3, dead(died)
. test drug2=drug3

```

```

1 if patient | Summary of Months to death or end of exp.
      died |      Mean   Std. Dev.   Freq.
-----+-----
      0 |   21.117647  10.652948    17
      1 |   12.419355   8.7512825   31
-----+-----
Total |      15.5   10.25629    48

```

```

Drug type | Summary of Months to death or end of exp.
(1=placebo) |      Mean   Std. Dev.   Freq.
-----+-----
      1 |   9.0526316   6.6204539   19
      2 |   14.5       7.2318739    6

```



3	21	10.620734	6
-----+			
Total	12.419355	8.7512825	31
Drug type	Summary of 1 if patient died		
(1=placebo)	Mean	Std. Dev.	Freq.
-----+			
1	.95	.2236068	20
2	.42857143	.51355259	14
3	.42857143	.51355259	14
-----+			
Total	.64583333	.48332111	48

```
Cox regression                                Number of obs =    48
                                                chi2(3)           =   36.52
Log Likelihood =-81.652567                    Prob > chi2       = 0.0000
```

Variable	Coefficient	Std. Error	t	Prob >  t	Mean
-----+					
studytim					15.5
died					.6458333
-----+					
age	.11184	.0365789	3.058	0.004	55.875
drug2	-1.71156	.4943639	-3.462	0.001	.2916667
drug3	-2.956384	.6557432	-4.508	0.000	.2916667
-----+					

( 1) drug2 - drug3 = 0.0, F(1, 45) = 3.31, Prob > F = 0.0757

17 个病人仍然存活，他们的生存时间自然要长。死亡的病人中，编号为 2 的药好于 1，而编号 3 的好于 2。似乎编号为 1 的药最差，使用模型  $h(t) = h_0(t) \exp(b_1 \text{age} + b_2 \text{drug}_2 + b_3 \text{drug}_3)$  并考虑到截尾，使用 `dead()` 选项，先对编号 2 和 3 的药生成指示变量，就可以使用 `cox` 命令了。对数似然值由 -99.911448，经四次迭代为 -81.652567。最后检验与第一种药是有区别的，第二、第三两种药是否有所不同。

`Survive.Kit` 是一组 Stata 用于生存分析的程序，使用命令 `run Survive.Kit` 装入。在 Stata 3.0 以后以 ADO 文件格式调用。`kapmeier` 和 `gwood` 用于绘制 Kaplan-Meier 生存曲线和给出根据 Greenwood 公式给出的置信带。`loglogs` 绘制  $\log(-\log(S(t)))$  对  $\log(t)$  的图，其中  $S(t)$  为由 Kaplan-Meier 积矩统计量定义的生存函数。若曲线为直线，则数据服从威布尔(Weibul)分布。`survsum` 和 `survcurv` 用于显示生存分析的综合统计量和生成生存变量。`logrank` 计算两组或多组 log-rank 统计量，用于比较两个或多个组的生存曲线。`mantel` 和 `wilcoxon` 计算两组时的 Mantel-Haenszel 检验及 Wilcoxon-Gehan 检验。另外，`coxhaz`、`coxbase` 和 `coxvar` 给出 `cox` 回归分析后基线风险函数、生存曲线和基线生存变量。这些命令通用的格式是：

<命令> 时间变量死亡指示变量[, by(分组变量)]

死亡指示变量取值为 1 时表示死亡，程序及结果如下：

```
. kapmeier studytim died
. kapmeier studytim died, by(drug)
. gwood studytim died
. gwood studytim died, by(drug)
. survsum studytim died
. survsum studytim died, by(drug)
. loglogs studytim died
. loglogs studytim died, by(drug)
. survcurv studytim died
. gen loglogs = log(-log(_surv))
. gen logt = log(studytim)
. regress loglogs logt
. logrank studytim died, by(drug)
```

Source	SS	df	MS	Number of obs =	48
Model	41.886338	1	41.886338	F( 1, 46) =	4439.69
Residual	.433987958	46	.009434521	Prob > F =	0.0000
Total	42.320326	47	.900432467	R-square =	0.9897
				Adj R-square =	0.9895
				Root MSE =	.09713

Variable	Coefficient	Std. Error	t	Prob >  t	Mean
loglogs					-.7396131
logt	1.075852	.0161464	66.631	0.000	2.449985
_cons	-3.375435	.0419694	-80.426	0.000	1

Group	Events	Predicted
1	19	7.2459559
2	6	8.1984653
3	6	15.555579

Chi2( 2 ) = 25.526241 , P = 2.864e-06

结果表明三组的确不同，P-值小于0.1%，注意到有19+6+6=31个死亡以生(标记为)预测值为31。对于对照组，(第一组1)，观察到19个死亡，但若所有病人具有相同的生存率。只会有7.25个死亡发生。

## §8.2.6 Stat.Kit

是一组程序，用于一系列统计检验。通过命令run Stat.Kit 调入。它提供了十五个新命令。

dbeta	计算影响统计量DF-Betas
genrank	产生变量的秩次，并考虑到相同秩次
genstd	产生标化变量(均值 0，方差1)
glogit	分组logit
gprobit	分组probit
ksmirnov	Kolmogorov-Smirnov 分布相等检验
kwallis	Kruskal-Wallis 单因素方差分析
means	算术平均、几何平均和调和平均
ranksum	Wilcoxon 秩和(Mann-Whitney 两样本) 统计量
regdw	回归并产出Durbin-Watson 统计量
signrank	Wilcoxon 配对符号秩次检验
signtest	配对检验中位数检验
spearman	Spearman 秩和相关系数
teststd	检验方差是否相同或者是否为某个常数
ttest	保种类型的t-检验

现要检验一种新的燃料添加剂的有效性，用12辆汽车实验，进行有无添加剂的里程的比较，结果如下：

	Without Treatment	With Treatment	Without Treatment	With Treatment
	20	24	18	17
	23	25	24	28
	21	21	20	24
	25	22	24	27
	18	23	23	21
	17	18	19	23

  

```
. summarize
```

Variable	Obs	Mean	Std. Dev.	Min	Max
mpg0	12	21	2.730301	17	25
mpg1	12	22.75	3.250874	17	28

  

```
. means
```

	Arithmetic		Geometric		Harmonic	
	Mean	Obs	Mean	Obs	Mean	Obs
mpg0	21	12	20.83629	12	20.67272	12

```
mpg1      |      22.75   12   22.52909   12   22.30015   12
```

```
. ttest mpg0=20
. ttest mpg0=mpg1, paired
. ttest mpg0=mpg1
. signtest mpg0=mpg1
. signrank mpg0=mpg1
```

以上命令进行均值是否20的检验，配对检验和成组t-检验。使用命令'ttest mpg0=mpg1, unequal'不做方差齐性的假设。其次用signtest命令检验中位数是否相同，运用Wilcoxon配对符号秩次检验检验分布是否相同。

接下来使用人口普查数据census.dta，检验结婚率，首先生成变量，用结婚数除以18岁以上的人数，然后考察结婚率与年龄的相关，可以使用原有的pearson相关命令或spearman命令。

```
. use census, clear
. generate mrgrate = marriage/pop18p
. summarize mrgrate
. correlate mrgrate medage
. spearman mrgrate medage
. summarize mrgrate, detail
```

## §8.3 高分辨统计制图

### §8.3.1 图形命令

Stata 1.0 的高分辨统计制图功能是靠Stata/GRAPH 完成的，Stata 2.0 的制图功能被集成在命令graph，其语法为：

[by varlist:] graph 变量列表[=权] [in 范围] [if 条件表达式] [, 选项] 或: graph using 文件名[文件名[.]] [, 选项]

使用graph命令可以绘制八种图形，即histogram(直方图)、matrix(二维散点图距阵，最多可指30个变量)、box(多达六个变量的Box-whisker图)、bar(直条图)、tway(二维散点图)、oneway(一维条形图，可与box结合使用)、star(星形图，最多有16个变量)和pie(圆图)，图形可以重显和并联。

```
graph using 文件名[文件名[.]] [, 选择项]
选项t|b|l|r1|2title("标题") margin(#) saving(文件名[,replace])
```

用例：

```
. graph x, saving(hist)
. graph y, saving(hist2)
. graph y x, saving(tway)
. graph y x, by(region) saving(many)
. graph using hist hist2 tway many
```

注意同时用using和saving()同样的文件名有时会破坏原有的文件。

共用的选择如：

1. saving(文件名[,replace]) 图形文件名, replace 指示覆盖原有文件。
2. by(变量名) 指示分组产生图形, 分组变量应按升序排列。
3. total 结合by() 指示产生整个数据的图形。
4. bsize(#) 指示分组变量标号的字体大小。
5. Rescale 仅结合by() 使用, 使每个分组量用不同的尺度。
6. title("标题") 在图形的下端加标题。多数不包含特殊字符的情况下, 引号可以省略。在图形的上下左右分别可以有两个标题, 简记t1()、b1() 等。
7. x|y|r|t<n>title("标题") 使用set textsize # 来控制标题文字的大小。
8. gap(#) 显示数轴时gap() 设定左标题和图轴数值的大小, 默认为8。
9. x|y|r|t|lable([#,...,#]) 和x|y|r|t|tick([#,...,#]) 与标题相仿, 用于指示指示图轴标号和图的标度。x|y|r|t|line([#,...,#]) 指示在图中加上纵横分隔线。x|y|r|scale(,#,#) 指示用新的较宽的尺度绘图。
10. noaxis 指示不显示图轴。
11. [no]border 用于控制图的边框。
12. log 指示直方图用对数尺度。在两维图中用rlog、ylog、xlog 来指示。
13. pen(#..#) 号码由一到九, Stata 用一号笔标记图的标号。
14. shading(#..#) 阴影深度, 号码越大越深, 用于直方图、直条图和圆图。
15. symbol(c..c) 指示twoway 和matrix 图中的符号, 可简写为s()。有O 大圆、o 小圆、S 大方块、d 小菱形、T 大三角、p 小加号[变量名] 把变量名用做标记、. 点、[\_n] 使用记录号、i 隐含。
16. trim(#) 当使用变量名做符号时, 图形中最多存放的字符数, 默认为8。名称"California", 因此graph y x, s(变量名) trim(2) 中符号变成为Ca。
17. psize(#) 指示[变量名] 符号的大小, 默认为100。
18. connect(c..c) 指示twoway 和matrix 中点间的连接方法, 可简写为c()。 . 表示不连, l 在点间画直线, m 连接median bands, s 表示用三次样条连接。
19. bands(#) 指示x-轴上bands 的数目, 此时计算x 与y 的中位数。
20. density(#) 指示在样条间的带子数, 默认为5。
21. bin(#) 指示区间数目, 默认为5。
22. freq 指示纵轴用频数单位代替分数单位。

23. `normal[(#,#)]` 指示在直方图上方画一个正态密度曲线。
24. `density(#)` 仅与`normal` 默认为100。
25. `jitter(#)` 增加一个随机噪声。`#` 表示噪声占图形区的比例, 最大为5。
26. `rescale` 指示每个y-变量分别规格化。
27. `rbox` 指示图上显示一个box 图。
28. `y|x|rreverse` 指示尺度是从高到底。
29. `jitter(#)` 指示# 个点在轴上显示出来。
30. `half` 指示仅用下三角阵。
31. `jitter(#)` 与`twoway` 相仿。
32. `[no]alt` 指示在箱式图下的标号是否要交错。
33. `vwidth` 产生不同宽度的箱式图, 图的宽度与观察数成比例。
34. `root` 指示`vwidth` 和`root` 选项后, 图的宽度与观察数目方根成正比。
35. `[no]alt` 与`box` 相仿。
36. `means` 指示用均值而非用变量的和规格化。
37. `stack` 指示图形是叠式的而非毗邻的。
38. `label(变量名)` 指示对每个记录用给定变量名的内容做标号。
39. `select(#,#)` 用于放大数据中某些星形图。

21-24用于直方图, 25-28用于二维图, 29用于一维图, 30-31用于矩阵图, 32 -34用于箱式图, 35-37用于条形图, 38和39用于星形图。

用例:

`graph y x =pop`, 绘x 与y 的图, 以圆的大小指示pop 的大小。

`graph y x =pop, psize(150)`, 与上同, 但所有符号增大50%。

`graph y x =pop, s([name])`, 使用变量name 的内容作标记, 字体大小与pop 成比例。

`graph y x =pop, symbol(.) jitter(4)` 随机地指示圆点表示pop 的大小。

`graph y hat x =pop, s(Oi) c(.1) sort` 绘制y hat 关于x 的图, 用直线连接各点。hat 是由`regress y x =pop` 回归并由`predict hat` 语句而来的预测值。

使用汽车数据的例子, 用Stata 绘制直方图、双向图、矩阵图、箱式图、星形图、直条图和圆图。

```
. graph mpg, histogram normal saving(hist)
. graph price mpg, twoway saving(twoway)
. graph price mpg weight length, matrix saving(matrix)
. graph weight, oneway by(foreign) total saving(onwway)
. graph weight, box by(foreign) total saving(box)
```

```
. graph price mpg displ wweight in 1/9, star saving(star)
. graph rep1-rep5, bar saving(bar)
. graph rep1-rep5, pie saving(pie)
. graph using hist twoway matrix oneway box star bar pie saving
(combined) title(The Combined Graphic Analysis)
```

### §8.3.2 图形打印

Stata 2.0 使用程序gphdot 与gphpen 进行图形硬拷贝。它们是独立的程序，虽然可以在Stata 下使用SHELL 来做，但最好在DOS 系统下来做。

gphdot 支持象素输出(pixel-oriented) 的设备如点阵打印机和HP laser 打印机。而gphpen 支持向量输出(vector-oriented) 设备如使用绘图笔的绘图仪和PostScript 打印机。

语法: gphdot|gphpen 文件名[/option /option ...]

文件名即图形(.gph) 文件，选项有:

/Lp|l (仅对gphdot) 指示portrait/landscape 状态。

/Dfilename 指示打印机描述文件，默认值是default.gdi/default.pen.

/Odevice: 输出设备。gphdot 通常送往并行口，如PRN:。/Ocom1 将送往

/Ofilename COM1:。/Omyfile.bin 输出至myfile.bin ( 可以用DOS 的COPY/B 命令送往打印机，因为文件是二进制格式)。gphpen 一般送往COM1:，但/Oprn: 则是送往PRN: /Omyfile.asc 是送往myfile.asc，除非是.pic 文件，gphpen 的输出总是ASCII 类型。用DOS 的COPY 或PRINT 命令可送往打印机。

/- 图形打印后不换页

/+ 开始打印时先换页

/C# 指示拷贝份数，不加指示时为1

/N 不打印Stata 标志

/R# 改变图形的大小，默认为/R100

/RX# 图形水平方向大小调整

/RY# 图形纵向调整

/S# 指示所有Stata 符号的大小，默认是/S100

/SO|S|T|o|d|p|. # 指示特定符号的大小，如/SO50

/I文件名对PostScript 使用ps.plf, 对.pic 文件用pic.plf。

/X (仅对gphpen) 使用绘图笔放慢速度以防墨溅到透明薄膜上。

/P#..# 改动默认绘图笔号/P123456789. P122 指示3 号笔同2号笔。

/T#..# 指示笔的厚度。默认为1。使用300-dpi 的设备时，指示为4可使图形好看一些。/T144 将使一号笔为1, 2 号笔与3 号笔为4。

打印上面的图形，在DOS 系统下，运行命令:

```
C>gphdot combined /r52 /Dhplphr
```

图形文件combined 将在HP laserJet+ 上印出。

### §8.3.3 记录文件输出

Stata 的运行记录文件虽然可以在DOS 下利用PRINT 命令印出，但使用系统提供的格式化程序FSL 并结合gdf 和profile.fsl 可以把Stata 的关键字醒目地显示出来。与其它软件相比，这一功能使人耳目一新。

运行方法: Dos: fsl i文件名[.log] [/选项]

Unix: fsl [-选项] i文件名[.log]

l设计名	格式化运行记录的设计名字
lfx	Epson FX, LX, RX
llq	Epson LQ, SQ
lmx	Epson MX
lhp	HP LaserJet
lhpfonds	HP LaserJet with prestige elite fonts
libmgr	IBM Graphics printer
lhispeed	IBM Mainframe
lnecp67	NEC P6 or P7
lnec8023	NEC 8023A-C
lokidata	Okidata 84, 92, 93, 182, 183
ltoshiba	Toshiba 3-in-1
lelite	Other, elite font or 90+ characters across
ldefault	Other, less than 90 characters across (default)
p文件名	包含设计的文件名, 默认为pprofile.fsl。
o文件名	输出文件名, 默认为o文件名.prn
h["]标题["]	页标题
u["]用户名["]	用户的名字
a["]帐号["]	帐号/代码
sc p	压缩输出类别
q	迅速执行

选项是:

### §8.3.4 graph.kit 与qc.kit

在stata 2.0中graph.kit与qc.kit定义一组绘图和质量控制图的命令, 用run 命令引用, 使用命令help graph.kit或help qc.kit 将获得它们用法的说明。在后期生动本中已经被ado文件取代。

grebar	error-bar chart
leverage	偏回归杠杆图
quantile	分位点图
qnorm	分位点—正态图
qqplot	分位点—分位点图
symplot	对称图

graph.kit 的命令有:

若没有图形显示设备, 则只能在图形打印设备上输出这些图形, 使用Sunview 下的Sun 时, 预先设置>window define tut'效果会好一些。使用命令query 可以得到计算机的配置情况。命令是:

```
. use auto, clear
. keep make price mpg weight displ foreign
. hilite mpg displ, hilite(foreign) ylabel xlabel
画一个两维散点图, 并且使部分数据醒目。
```



可以使用`regress mpg displ weight foreign` 或 `regress mpg foreign weight displ` 命令进行一般回归分析, 右边变量的次序与分析无关, 对应于车重的偏回归杠杆图则不然, 必需使`weight` 成为方程右边第一个变量。相应于指示变量`foreign`, 可以使用命令

```
. leverage mpg foreign displ weight
```

有时我们对一个变量的分布是否对称感兴趣, 可以使用对称图示法。

汽车数据中最高的车价为\$15,906, 最低的为\$3,291, 现在将它俩与数据中典型的车型比较, 若分布对称, 价格的差别应是雷同的。车价的中位数为\$5,006.50, 因此最贵的车多花\$10,899.50 而最便宜的车少花\$1,715.50, 由此看它们是对称的吗? 对其它的数据重复此做法, 就是对称图的思想, 现只消用命令。

```
. symplot price
```

qc.Kit 的命令是: `cchart`, `pchart`, `plotebar`, `shewhart`。现用Acheson J. Duncan 著Quality Control and Industrial Statistics 中的数据, 由每天流出的50个railway frames 样本所预计的railway frames数目, 这些数据是五月份头28天的数据。

```
. describe
```

```
. summarize rejects, detail
```

```
. pchart rejects day ssize, ylabel xlabel
```

```
. pchart rejects day ssize, ylabel xlabel
```

```
. pchart rejects day ssize, xlabel ylabel stabilized title(May Production)
```

为演示命令`cchart`, 使用新的数据, 它记录了每组5台, 共25组收音机的发现问题的数目。

```
. cchart defects sample, ylabel xlabel title(c-Chart for Radio Subassemblies)
```

```
. list
```

	date	mean	std
1.	8	192.2194	3.937047
2.	9	192.6444	2.833564
3.	10	192.3667	4.578077
4.	13	194.7625	3.250883
5.	14	192.6889	2.889829
6.	15	195.0182	1.730357
7.	16	193.4028	2.61576

```
. plotebar mean std date, yline(195) title(Weight Variation) ylabel xlabel
```

最后是cusum 图, 用`graph` 命令就可以了。

```
. graph sum top bot unit, yline(0) ylabel xlabel c(.ll) s(oii)
```

