

STATA基本代码

2019年9月11日

一、基本命令语法

1. 语句结构

[by varlist:] **command** [varlist] [=exp] [if exp] [in range] [weight] [, options]

2. 各结构含义

[by varlist:]: 按照某已排序的变量分类执行指令

command: 指令内容

[varlist]: 执行指令的变量

[=exp]: 赋值

[if exp]: 条件语句

[in range]: 指定某一个范围

[weight]: 用某变量加权

[, options]: 可选项——命令的补充说明

3. 一些常用操作

以系统自带数据为例 (sysuse auto)

1. 分类操作[by varlist:]

sort foreign //将分类标准进行排序

by foreign: **sum** price weight //根据排序过的变量分类描述统计

2. 赋值及其运算[=exp]

gen nprice = price + 10 //新生成变量nprice

replace nprice = price - 10 //将nprice用新的表达式覆盖

list price nprice //列出数据以对比

3. 条件语句[if exp]

list make price if foreign == 0 //只看国产车的品牌和价格

list make price if foreign==0 & price > 10000 //只看价格超过一万元的进口车

list make price if foreign==0 | price > 10000 //只看价格超过一万元的车或者进口车

4. 范围筛选[in range]

sort price //将价格进行排序

sum price in 1/10 //对前十个价格进行描述性统计

5. 加权[weight]

sum price [weight=length] //按照车长对价格进行加权描述性统计

6. 可选项[, options]

sum price, detail //更细致的描述性统计

list price in 1/30, sep(10) //每10个观测值之间加一个横线

list price, nohead //不要表头

二、数据分析过程

0.数据清理

1.clear

功能：清空之前的所有数据

写法：**clear**

2.rename

功能：重命名变量

写法：**rename** fe1013ge price //将原本名为fe1013ge改名为price

3.lable

功能：给变量加上标签

写法：**label** var p "price" //注意变量名前面要加var，标签名称要打引号

4.drop

功能：舍弃观测值

写法：**drop** if (p == . | p <= 0) //舍弃价格为空值或者为负值的观测值

1.描述性统计

1.list

功能：将观测值列出

写法：

list price in 1/30 //将前30个price列出

list price weight //将price和weight列出

示意图：

	price	weight
1.	4,099	2,930
2.	4,749	3,350
3.	3,799	2,640
4.	4,816	3,250
5.	7,827	4,080
6.	5,788	3,670
7.	4,453	2,230
8.	5,189	3,280
9.	10,372	3,880
10.	4,082	3,400
11.	11,385	4,330
12.	14,500	3,900
13.	15,906	4,290
14.	3,299	2,110
15.	5,705	3,690
16.	4,504	3,180
17.	5,104	3,220
18.	3,667	2,750
19.	3,955	3,430

—more—

2.tabstat

功能：基本描述性统计

- 1) 必须带上变量名
- 2) 什么都不指定的时候, 只显示平均数
- 3) 加上stat()可选项之后, 可以指定想要显示的指标

写法:

tabstat price weight //什么可选都没有, 就只列出平均值

tabstat price length, stat (max min mean p50 sd n) //指定列出最大值、最小值、平均、中位数、标准差、有效观测数

示意图:

- 1) 什么都不指定

stats	price	weight
mean	6165.257	3019.459

- 2) 指定stat

stats	price	length
max	15906	233
min	3291	142
mean	6165.257	187.9324
p50	5006.5	192.5
sd	2949.496	22.26634
N	74	74

3.summarize

功能: 一般/细致的描述性统计

- 1) 什么都不加的时候 (甚至变量名也不带), 默认显示全部变量的观测数、平均数、标准差、最小值和最大值
- 2) 加上可选项detail之后, 在之前的基础上加上了频率区间、总权重、方差、偏度 (skewness) 和峰度 (kurtosis)
- 3) 常用的统计量缩写

表 1: 常用统计量表

统计量	含义	统计量	含义
mean	均值	semmean	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		

写法:

- `sum` //全部变量的常用统计量
- `sum p` //只对price做描述性统计
- `sum p, detail` //对p做细致的描述性统计

示意图:

1) 什么都不加

Variable	Obs	Mean	Std. Dev.	Min	Max
make	0				
price	74	6165.257	2949.496	3291	15906
mpg	74	21.2973	5.785503	12	41
rep78	69	3.405797	.9899323	1	5
headroom	74	2.993243	.8459948	1.5	5
trunk	74	13.75676	4.277404	5	23
weight	74	3019.459	777.1936	1760	4840
length	74	187.9324	22.26634	142	233
turn	74	39.64865	4.399354	31	51
displacement	74	197.2973	91.83722	79	425
gear_ratio	74	3.014865	.4562871	2.19	3.89
foreign	74	.2972973	.4601885	0	1

2) 加上detail

Price					
Percentiles	Smallest				
1%	3291	3291			
5%	3748	3299			
10%	3895	3667	Obs		74
25%	4195	3748	Sum of Wgt.		74
50%	5006.5		Mean		6165.257
		Largest	Std. Dev.		2949.496
75%	6342	13466			
90%	11385	13594	Variance		8699526
95%	13466	14500	Skewness		1.653434
99%	15906	15906	Kurtosis		4.819188

4.pwcorr

功能: 写出相关系数矩阵

写法:

- `pwcorr` //所有变量的相关系数矩阵
- `pwcorr p w` //价格和重量的相关系数矩阵
- `pwcorr, sig star (0.05)` //显著性水平为0.05标星的相关系数矩阵

示意图:

1) 普通的

(make ignored because string variable)

	p	mpg	rep78	headroom	trunk	weight	length
p	1.0000						
mpg	-0.4686	1.0000					
rep78	0.0066	0.4023	1.0000				
headroom	0.1145	-0.4138	-0.1480	1.0000			
trunk	0.3143	-0.5816	-0.1572	0.6620	1.0000		
weight	0.5386	-0.8072	-0.4003	0.4835	0.6722	1.0000	
length	0.4318	-0.7958	-0.3606	0.5163	0.7266	0.9460	1.0000
turn	0.3096	-0.7192	-0.4961	0.4245	0.6011	0.8574	0.8643
displacement	0.4949	-0.7056	-0.4119	0.4745	0.6086	0.8949	0.8351
gear_ratio	-0.3137	0.6162	0.4103	-0.3779	-0.5087	-0.7593	-0.6964
foreign	0.0487	0.3934	0.5922	-0.2939	-0.3594	-0.5928	-0.5702
	turn	displa	t gear_r	o foreign			
turn	1.0000						
displacement	0.7768	1.0000					
gear_ratio	-0.6763	-0.8289	1.0000				
foreign	-0.6311	-0.6138	0.7067	1.0000			

2) 带显著性水平的

(make ignored because string variable)

	p	mpg	rep78	headroom	trunk	weight	length
p	1.0000						
mpg	-0.4686*	1.0000					
rep78	0.0066	0.4023*	1.0000				
headroom	0.1145	-0.4138*	-0.1480	1.0000			
trunk	0.3143*	-0.5816*	-0.1572	0.6620*	1.0000		
weight	0.5386*	-0.8072*	-0.4003*	0.4835*	0.6722*	1.0000	
length	0.4318*	-0.7958*	-0.3606*	0.5163*	0.7266*	0.9460*	1.0000
turn	0.3096*	-0.7192*	-0.4961*	0.4245*	0.6011*	0.8574*	0.8643*

5.其他描述性统计

1.describe

功能：总体展示数据情况

(变量数、观测数、总空间、数据类型、存储格式、变量名、标签)

- 不加变量名

示意图：

```

Contains data from C:\Program Files (x86)\Stata14\ado\base/a/auto.dta
  obs:          74                1978 Automobile Data
  vars:         12                13 Apr 2014 17:45
  size:        3,182             (_dta has notes)

```

variable name	storage type	display format	value label	variable label
make	str18	%-18s		Make and Model
price	int	%8.0gc		Price
mpg	int	%8.0g		Mileage (mpg)
rep78	int	%8.0g		Repair Record 1978
headroom	float	%6.1f		Headroom (in.)
trunk	int	%8.0g		Trunk space (cu. ft.)
weight	int	%8.0gc		Weight (lbs.)
length	int	%8.0g		Length (in.)
turn	int	%8.0g		Turn Circle (ft.)
displacement	int	%8.0g		Displacement (cu. in.)
gear_ratio	float	%6.2f		Gear Ratio
foreign	byte	%8.0g	origin	Car type

Sorted by: foreign

2.codebook

功能：展示数据库中每个变量情况

- 可加变量名

codebook price

示意图：

```

make                                                                    Make and Model
-----
type:  string (str18), but longest is str17
unique values:  74                missing "":  0/74
examples:  "Cad. Deville"
           "Dodge Magnum"
           "Merc. XR-7"
           "Pont. Catalina"
warning:  variable has embedded blanks

price                                                                    Price
-----
type:  numeric (int)
range:  [3291,15906]
unique values:  74                units:  1
                                           missing .:  0/74
mean:  6165.26
--more--

```

3.count

功能：报告有多少个观测值

- 不加变量名

by foreign: **count**

示意图：（加上by之后）

```
. by foreign: count  
  
-----  
-> foreign = Domestic  
    52  
-----  
-> foreign = Foreign  
    22
```

4.browse

功能：浏览数据

- 可加变量名

br price in 1/30

示意图：

	make	price	mpg	rep70	headroom	trunk	weight	length	turn	displacement	gear_ratio	A
1	Herc. Amptur	3,191	29	3	3.5	17	2,030	195	43	140	3.00	
2	Chev. Chevette	3,199	29	3	2.5	9	2,110	163	34	231	2.93	
3	Chev. Monza	3,667	24	2	2.0	7	2,750	179	40	151	2.73	
4	Toyota Corolla	3,740	31	5	3.0	9	2,200	165	35	97	3.21	
5	Subaru	3,790	35	5	2.5	11	2,050	164	38	97	3.01	
6	AMC Spirit	3,799	22	.	3.0	12	2,440	160	33	121	3.00	
7	Herc. Bobcat	3,829	22	4	3.0	9	2,900	169	39	140	2.73	
8	Renault Le Car	3,895	24	3	3.0	10	1,930	142	34	79	3.72	
9	Chev. Nova	3,955	19	3	3.0	13	3,430	187	43	250	2.54	
10	Dodge Colt	3,964	30	5	2.0	8	2,120	163	35	91	3.04	
11	Honda GC	3,990	30	4	3.0	11	1,990	194	33	98	3.73	
12	Dodge Diplomat	4,010	18	3	4.0	17	3,400	206	44	314	2.47	
13	Flym. Volare	4,040	18	3	3.0	14	3,330	201	44	228	3.23	
14	Daik. Skylark	4,082	19	3	3.0	13	3,490	200	42	231	3.08	
15	AMC Concord	4,099	22	3	2.5	11	2,930	184	40	121	3.00	
16	Ford. Mustang	4,172	24	2	2.0	7	2,440	179	41	151	2.73	
17	Olds Omega	4,181	19	3	4.5	14	3,370	200	43	231	3.00	
18	Ford Mustang	4,187	21	3	2.0	10	2,460	179	43	140	3.00	
19	Olds starfire	4,196	24	1	2.0	10	2,730	160	40	151	2.73	
20	Fiat Strada	4,196	21	3	2.5	14	2,130	161	36	101	3.27	
21	Ford Fiesta	4,199	28	4	2.5	9	1,900	147	33	98	3.15	
22	Ford. Pionix	4,424	19	.	3.5	12	3,420	203	43	231	3.00	
23	Flym. Champ	4,425	24	5	2.5	11	1,900	157	37	94	2.97	
24	Daik. Opel	4,453	26	.	3.0	10	2,230	170	34	304	2.67	
25	Flym. Horizon	4,462	25	3	4.0	17	2,200	165	38	105	3.37	
26	Honda Civic	4,499	28	4	2.5	5	1,760	149	34	91	3.30	
27	Chev. Malibu	4,504	22	3	3.5	17	3,130	193	31	208	2.73	
28	Herc. Monarch	4,514	18	3	3.0	15	3,370	190	41	250	2.43	
29	Saturn I10	4,569	35	5	2.0	9	2,020	165	32	95	3.70	
30	Flym. Arrow	4,647	29	3	2.0	11	3,200	170	37	158	3.05	

5.tabulate

功能：呈现频率分布

- 加一个变量名

`tab p`

示意图：

price	Freq.	Percent	Cum.
3,291	1	1.35	1.35
3,299	1	1.35	2.70
3,667	1	1.35	4.05
3,748	1	1.35	5.41
3,798	1	1.35	6.76
3,799	1	1.35	8.11
3,829	1	1.35	9.46
3,895	1	1.35	10.81
3,955	1	1.35	12.16
3,984	1	1.35	13.51
3,995	1	1.35	14.86
4,010	1	1.35	16.22
4,060	1	1.35	17.57
4,082	1	1.35	18.92
4,099	1	1.35	20.27
4,172	1	1.35	21.62
4,181	1	1.35	22.97
4,187	1	1.35	24.32
4,195	1	1.35	25.68
4,296	1	1.35	27.03
4,389	1	1.35	28.38
4,424	1	1.35	29.73
4,425	1	1.35	31.08

—more—

6.table

功能：变量联列表

- 加两个变量名

`tab pw`

示意图：

price	Weight (lbs.)					Total
	1,760	1,800	1,830	1,930	1,980	
3,291	0	0	0	0	0	1
3,299	0	0	0	0	0	1
3,667	0	0	0	0	0	1
3,748	0	0	0	0	0	1
3,798	0	0	0	0	0	1
3,799	0	0	0	0	0	1
3,829	0	0	0	0	0	1
3,895	0	0	1	0	0	1
3,955	0	0	0	0	0	1
3,984	0	0	0	0	0	1
3,995	0	0	0	0	1	1
4,010	0	0	0	0	0	1
4,060	0	0	0	0	0	1
4,082	0	0	0	0	0	1
4,099	0	0	0	0	0	1
4,172	0	0	0	0	0	1
4,181	0	0	0	0	0	1
4,187	0	0	0	0	0	1
4,195	0	0	0	0	0	1
4,296	0	0	0	0	0	1
4,389	0	1	0	0	0	1
4,424	0	0	0	0	0	1

—more—

2. 回归分析

1. reg

功能：普通OLS回归，报告系数、ANOVA分析、置信区间、R方、调整R方、F值等

- 加一个或多个变量名

```
reg p w l
```

示意图：

Source	SS	df	MS	Number of obs	=	74
Model	220725280	2	110362640	F(2, 71)	=	18.91
Residual	414340116	71	5835776.28	Prob > F	=	0.0000
Total	635065396	73	8699525.97	R-squared	=	0.3476
				Adj R-squared	=	0.3292
				Root MSE	=	2415.7

price	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
weight	4.699065	1.122339	4.19	0.000	2.461184	6.936946
length	-97.96031	39.1746	-2.50	0.015	-176.0722	-19.84838
_cons	10386.54	4308.159	2.41	0.019	1796.316	18976.76

3. 结果报告

1. logout

功能: 输出各种图表

写法:

1) 一般形式

`logout, save(文件名不含扩展) 文件形式(word excel等) [replace(覆盖原有同名文件)] :`
command

2)

`logout, save(test) excel replace: pwcorr p w l, sig star(0.05) //`输出以test为名的excel文件,
内容是p、w、l表明显著性的相关系数矩阵

示意图:

	price	weight	length	
price	1			
weight	0.5386*	1		
length	0.4318*	0.9460*	1	
	0.000100	0		

2. esttab

功能: 输出最小二乘回归表

写法:

`reg p mpg //`做第一个最小二乘回归est

`store m1 //`把它存储下来记为m1 `reg p`

`mpg w l //`做第二个最小二乘回归est

`store m2 //`把它存储下来记为m2

`esttab m1 m2 using reg1.rtf, r2 ar2 se replace nogap`

//一般形式为: `esttab 回归名1 回归名2 using 文件名加扩展名, r2 ar2 se` (展示r2调整r2并在括号里写标准差) [replace] [nogap] (不要空行)

示意图:

	(1)	(2)
	price	price
mpg	-238.9*** (-4.50)	-86.79 (-1.03)
weight		4.365*** (3.74)
length		-104.9* (-2.64)
cons	11253.1*** (9.61)	14542.4* (2.47)
N	74	74

t statistics in parentheses
* p < 0.05, ** p < 0.01, *** p < 0.001

3.outreg2

功能：输出最小二乘回归表

写法：

```
reg p mpg //做第一个最小二乘回归
```

```
est store m1 //把它存储下来记为m1
```

```
reg p mpg w l //做第二个最小二乘回归
```

```
est store m2 //把它存储下来记为m2
```

```
outreg2 using reg1.xls
```

```
//一般形式为： outreg2 using 文件名加扩展名, [replace]
```

3.示意图

	(1)	(2)
VARIABLES	price	price
mpg	-86.79 (83.94)	-238.9*** (53.08)
weight	4.365*** (1.167)	
length	-104.9** (39.72)	
Constant	14,542** (5,891)	11,253*** (1,171)
Observations	74	74
R-squared	0.357	0.220
Standard errors in parentheses		
*** p<0.01, ** p<0.05, * p<0.1		

4.统计制图

1.散点图-拟合直线、折线

指令:

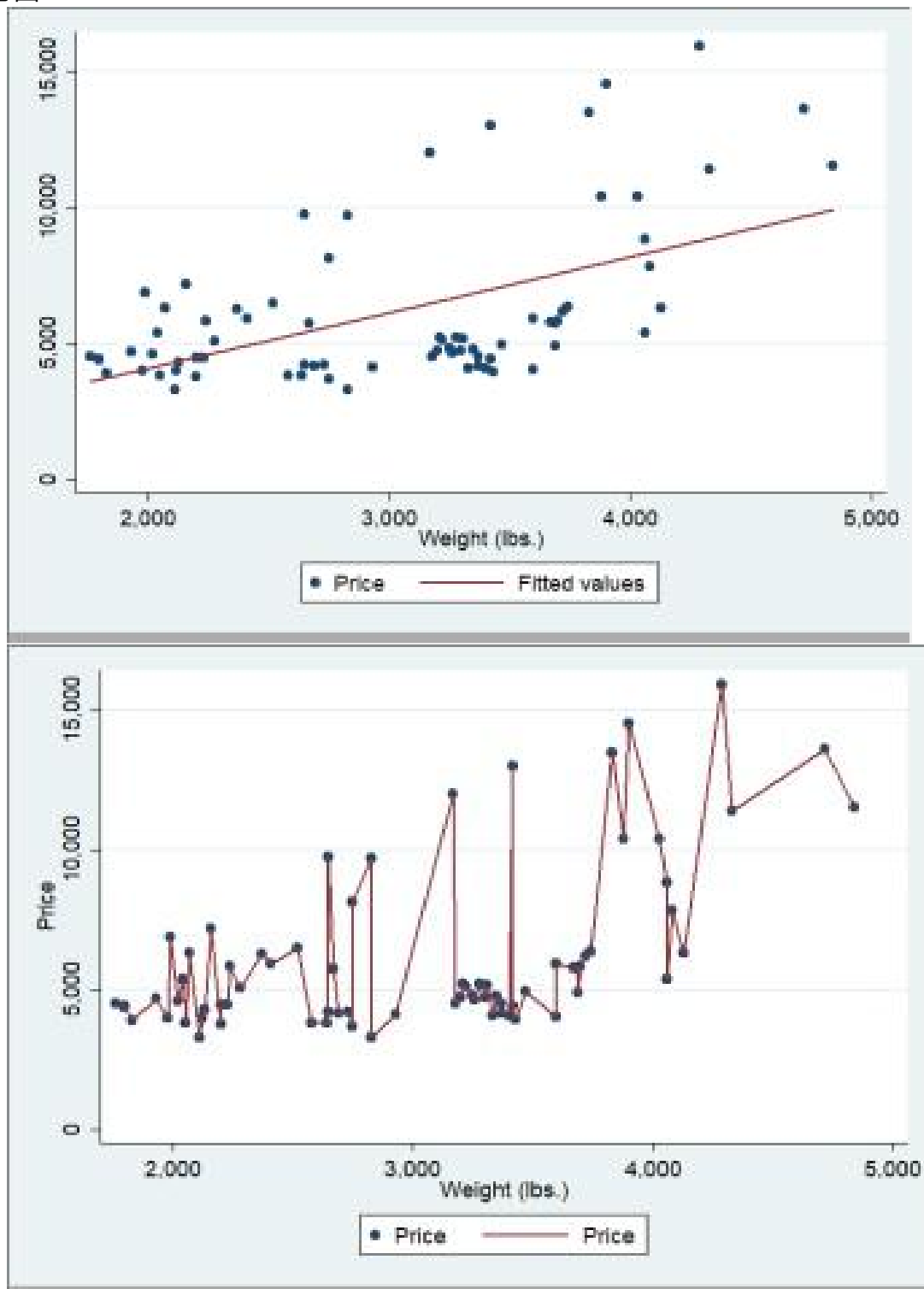
`sort w` //先将横坐标排序

`scatter p w` //画出散点图

`scatter p w || lfit p w` //画出散点图与拟合直线

`scatter p w || line p w` //画出散点图与折线

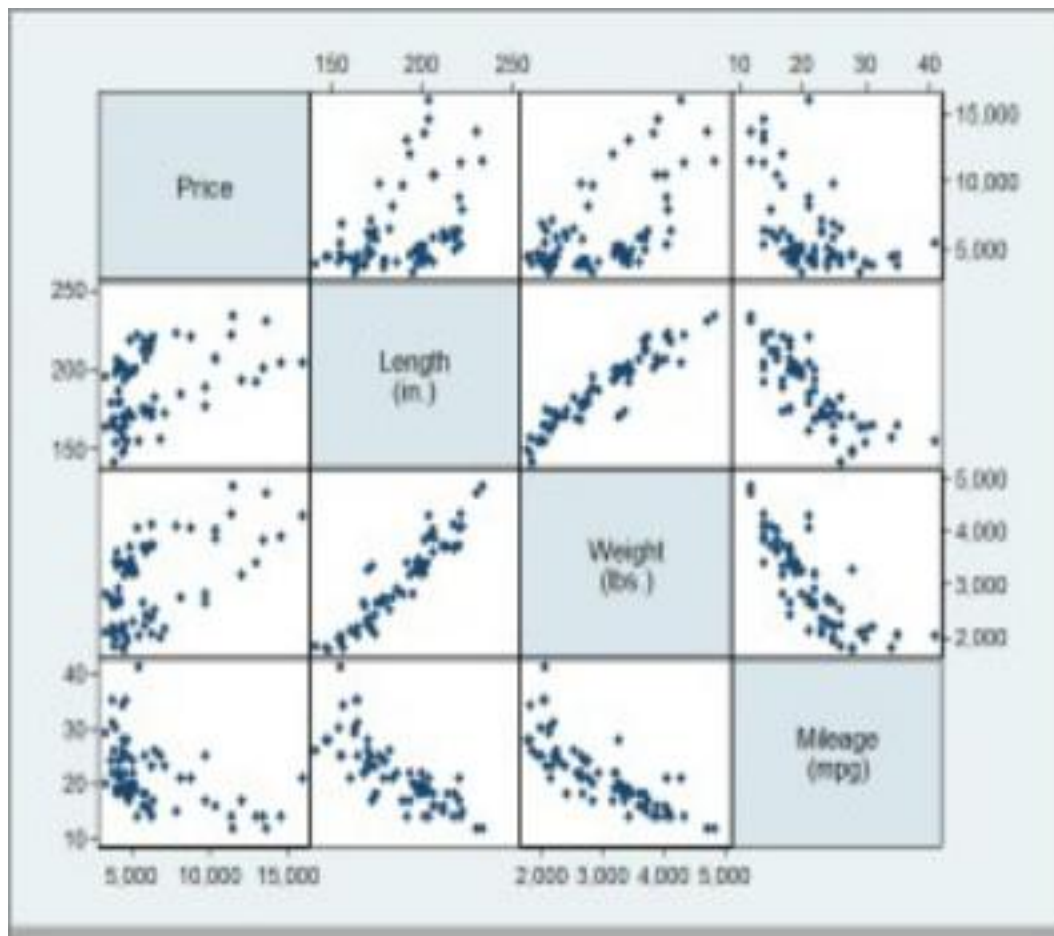
示意图:



2.相关矩阵图

指令：`graph matrix p l w mpg //`画出这四个变量的相关矩阵图

示意图：



3.条形图、饼形图、点阵图

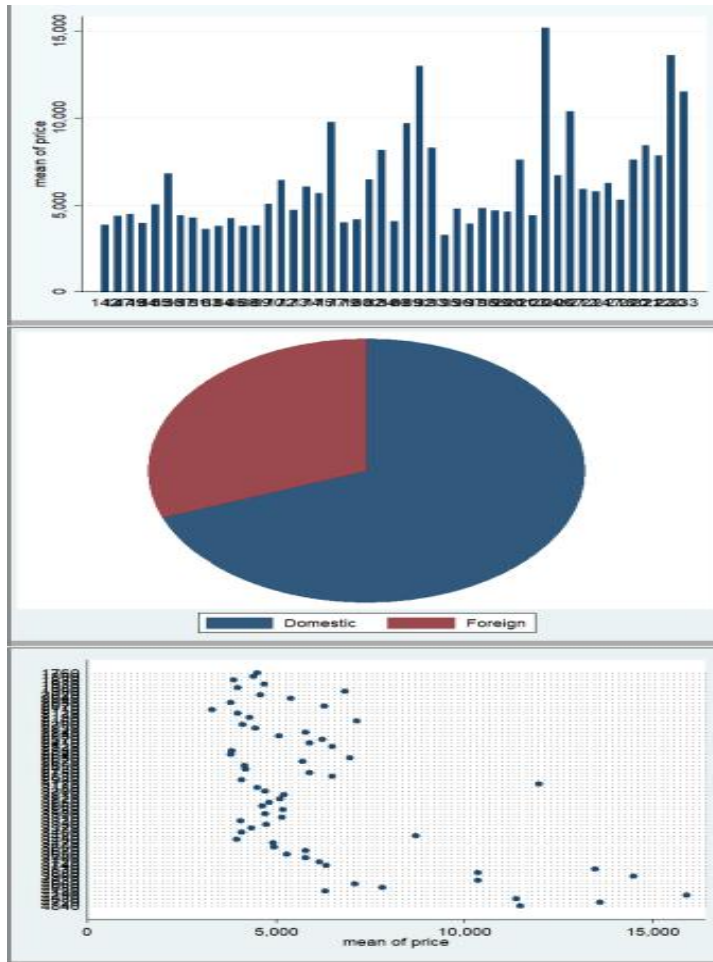
指令:

`graph bar price, over(foreign)` //画 (均值) 柱状图

`graph pie price, over(foreign)` //画饼状图

`graph dot price, over(weight)` //画 (均值) 点阵图.

示意图:



4.直方图

指令:

`histogram mpg, discrete frequency normal` //画出mpg的直方图, 并做出正态曲线

`histogram q, width(1000) frequency` //指定直方图宽度是1000

示意图:

