Stata pour débutant

Ibrahima TALL March 26, 2025

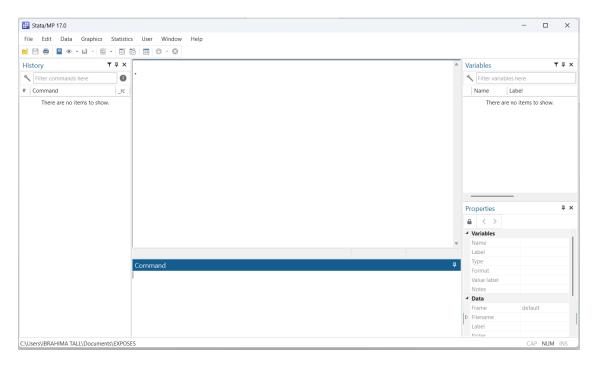
Contents

1	Introduction du logiciel	2
	1.1 Fenêtres de l'interface	2
	1.2 Commande stata	2
	1.3 Calculatrice	2
	1.4 Commandes système	3
2	Importation et exportation de données	5
	2.1 Importation de données	5
	2.2 Exportation de données	6
3	Exploitation de la base	6
	3.1 Observation de la base	6
	3.2 Informations sur les données	8
	3.3 Rangement de la base	9
	3.4 Commandes logique	10
	3.5 Doublons sur les observations	10
	3.6 Recherche de variables	11
	3.7 Statistiques usuelles	12
4	Tableaux de statistiques	14
	4.1 Tableaux univariés	14
	4.2 Tableaux bivariés	15
	4.3 Tableaux de statistiques	17
	4.4 Tableaux généralisés	18
5	Création et modification de variables	19
6	Création de variables	19
	6.1 Étiquettes (labels) de variables	19
	6.2 Modifier une variable	
	6.3 Transformation en variables catégorielles	20
	6.4 Étiquettes des valeurs	
	6.5 Transformation texte et numérique	
	6.6 Suppression de variables et observations	
7	Illustration graphique	21

1 Introduction du logiciel

Un logiciel commercial de traitement statistique et de modélisation économétrique. La première version est sortie en Janvier 1985 et à la date d'écriture de ces lignes, le logiciel en est à sa version 18. Dans ce tutoriel, c'est la version 17 qui est utilisée.

1.1 Fenêtres de l'interface



1.2 Commande stata

```
help compare

search tabout

ssc install gsample
```

checking gsample consistency and verifying not already installed... all files already exist and are up to date.

```
net install gsample
```

checking gsample consistency and verifying not already installed... all files already exist and are up to date.

1.3 Calculatrice

```
display as txt "La somme est de S = " as res 1+6
```

La somme est de S = 7

```
display as res 7-5
2
display 2*7
14
display 17/3
5.6666667
display int(17/3)
5
display mod(17,3)
2
display 2^3
8
display exp(1)
2.7182818
display sin(_pi/2)
1
display comb(10,2)
45
mata: factorial(3)
  6
1.4 Commandes système
```

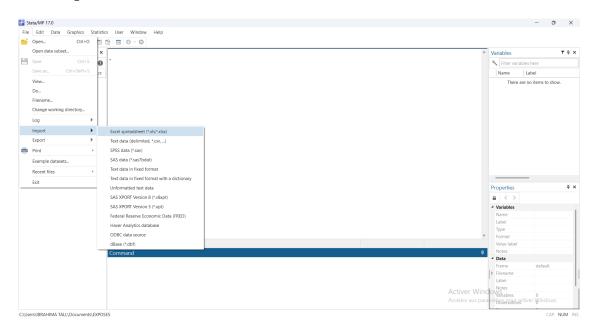
```
ls *.png
  10.3k 10/05/24 1:41 MPN.png
findfile MPN.png
./MPN.png
copy MPN.png PMN_new.png
ls stata*
file not found
rm PMN_new.png // erase
sysdir
   STATA: C:\Program Files\Stata17\
   BASE: C:\Program Files\Stata17\ado\base\
   SITE: C:\Program Files\Stata17\ado\site\
   PLUS: C:\Users\IBRAHIMA TALL\ado\plus\
PERSONAL: C:\Users\IBRAHIMA TALL\ado\personal\
OLDPLACE: c:\ado\
pwd
C:\Users\IBRAHIMA TALL\Documents
cd ..
C:\Users\IBRAHIMA TALL
cd "C:\Users\IBRAHIMA TALL\Documents"
C:\Users\IBRAHIMA TALL\Documents
mkdir EXPOSESR
mkdir mondoc
rmdir mondoc
cd ./EXPOSES
C:\Users\IBRAHIMA TALL\Documents\EXPOSES
which display
```

built-in command: display

2 Importation et exportation de données

Il est possible d'importer ou d'exporter des données via l'interface ou d'utiliser des commandes. L'interface est plus comode lorsque l'emplacement du fichier n'est pas précis. Dans le cas où le dossier de travail est le répertoire courant, la commande est seulement suivi du nom du fichier.

2.1 Importation de données

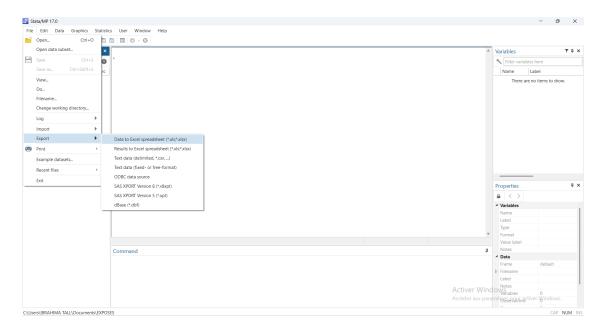


Les modes d'importation dépendent des types de fichiers de données: 1. fichier de type stata (.dta): pour ce dernier type, les données peuvent être sur disque ou dans le système de stata: - en mémoire: use ; - en système: sysuse dir (la liste) et sysuse; - en ligne: webuse. 1. fichier de type texte: import excel ; 2. fichier de type excel: import delimited ; 3. fichier de type SPSS (.sav): import spss.

```
sysuse dir
 auto.dta
                  census.dta
                                  network1.dta
                                                  surface.dta
 auto16.dta
                 citytemp.dta
                                  network1a.dta
                                                  tsline1.dta
 auto2.dta
                 citytemp4.dta
                                  nlsw88.dta
                                                  tsline2.dta
                 educ99gdp.dta
 autornd.dta
                                  nlswide1.dta
                                                  uslifeexp.dta
 bplong.dta
                  gnp96.dta
                                  pop2000.dta
                                                  uslifeexp2.dta
 bpwide.dta
                 lifeexp.dta
                                  sandstone.dta
                                                  voter.dta
 cancer.dta
                 mabase.dta
                                  sp500.dta
                                                  xtline1.dta
sysuse auto, clear
```

(1978 automobile data)

2.2 Exportation de données



L'exportation est faite en remplaçant export à la place de import. L'exportation au format .sav (SPSS) n'est pas disponible dans la version 17. Pour les données de type stata, l'exportation correspond à une sauvegarde avec la commande save.

```
rm mabase.dta
save mabase, replace
```

(file mabase.dta not found)
file mabase.dta saved

3 Exploitation de la base

```
use mabase, clear
```

(1978 automobile data)

price

mpg

Les commentaires sont explimés par l'asterix (*), le double (//) et triple (///) slash.

3.1 Observation de la base

headroom

trunk

```
* Noms de variables
describe, simple
make rep78 weight displacement
```

gear_ratio

foreign

length

turn

* Description de la base describe

Contains data from mabase.dta

Observations: 74 1978 automobile data
Variables: 12 26 Mar 2025 22:23
(_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 origin

Sorted by: foreign

codebook price

price

Type: Numeric (int)

Range: [3291,15906] Units: 1
Unique values: 74 Missing .: 0/74

Mean: 6165.26 Std. dev.: 2949.5

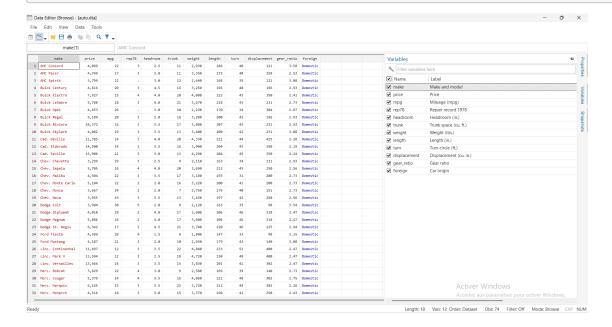
Percentiles: 10% 25% 50% 75% 90%

3895 4195 5006.5 6342 11385

list price mpg rep78 headroom trunk weight length in 1/7

	+			. – – – – – – – – –			+
	price	mpg	rep78	headroom	trunk	weight	length
1.	4,099	22	3	2.5	11	2,930	186
2.	4,749	17	3	3.0	11	3,350	173
3.	3,799	22	•	3.0	12	2,640	168
4.	4,816	20	3	4.5	16	3,250	196
5.	7,827	15	4	4.0	20	4,080	222
6.	5,788	18	3	4.0	21	3,670	218
7.	4,453	26	•	3.0	10	2,230	170
	+						+

browse // Ou edit



3.2 Informations sur les données

notes list

_dta:

1. From Consumer Reports with permission

notes: Les voitures américaines notes make: La marque et la serie de la voiture notes

_dta:

- 1. From Consumer Reports with permission
- 2. Les voitures américaines

make:

1. La marque et la serie de la voiture

notes replace _dta in 2: Les voitures d'occasion

(note 2 for _dta replaced)

notes _dta

_dta:

- 1. From Consumer Reports with permission
- 2. Les voitures d'occasion

notes search voiture

_dta:

2. Les voitures d'occasion

make:

1. La marque et la serie de la voiture

notes drop _dta in 2

(1 note dropped)

notes list

_dta:

1. From Consumer Reports with permission

make:

1. La marque et la serie de la voiture

3.3 Rangement de la base

sort make

gsort foreign -price

order turn foreign, after(make)

order foreign, last

3.4 Commandes logique

```
assert inrange(price,0, 100000)
```

isid make

3.5 Doublons sur les observations

```
* Ajout d'une ligne
set obs 75
```

Number of observations ($_{\rm N}$) was 74, now 75.

```
insobs 2, before(20)
```

(2 observations added)

```
expand 2 in 66/74
```

(9 observations created)

```
duplicates report make
```

Duplicates in terms of make

(Copies	Observations	Surplus
	1 2 3	65 18	 (9

```
duplicates list make
```

 ${\tt Duplicates} \ {\tt in} \ {\tt terms} \ {\tt of} \ {\tt make}$

Gr	oup	Obs	make	
	1	20		
ļ	1	21		
1	1	77 70	Do +	010
l I	2 2	70 82	Datsun Datsun	
i	3	68	Datsun	510
	3	80	Datsun	510
+				+

```
duplicates tag make, generate(repeted)
```

Duplicates in terms of make

```
list make price foreign if repeted == 1
```

	+		+
	make 	price	foreign
66.	 Toyota Corona	5,719	Foreign
67.	VW Diesel	5,397	Foreign
68.	Datsun 510	5,079	Foreign
69.	VW Rabbit	4,697	Foreign
70.	Datsun 210	4,589	Foreign
71.	Honda Civic	4,499	Foreign
72.	Fiat Strada	4,296	Foreign
73.	Mazda GLC	3,995	Foreign
74.	Renault Le Car	3,895	Foreign
78.	Toyota Corona	5,719	Foreign
79.	VW Diesel	5,397	Foreign
80.	Datsun 510	5,079	Foreign
81.	VW Rabbit	4,697	Foreign
82.	Datsun 210	4,589	Foreign
83.	Honda Civic	4,499	Foreign
84.	Fiat Strada	4,296	Foreign
85.	Mazda GLC	3,995	Foreign
86.	Renault Le Car	3,895	Foreign
	+		+

```
duplicates drop make, force
```

Duplicates in terms of make

(11 observations deleted)

3.6 Recherche de variables

```
lookfor "in."
```

Variable	Storage	Display	Value	
name	type	format	label	Variable label
headroom	float	%6.1f		Headroom (in.)
length	int	% 8.0g		Length (in.)
displacement	int	%8.0g		Displacement (cu. in.)

```
ds, has(vallabel origin)
foreign
ds, has(varlabel *in.*)
headroom
             length
                         displacement
ds, not(type numeric)
make
3.7 Statistiques usuelles
count if price <= 5000</pre>
 37
by foreign, sort: count if price <= 5000
-> foreign = Domestic
-> foreign = Foreign
\rightarrow foreign = .
inspect price
price: Price
                                            Number of observations
                                        Total
                                                   Integers Nonintegers
                          Negative
                          Zero
                          Positive
                                           74
                                                       74
                                    -----
                          Total
                                            74
                                                       74
      # . . .
                          Missing
                                            75
  (74 unique values)
summarize weight
   Variable
                 Obs Mean Std. dev.
                   74 3019.459 777.1936 1760
     weight |
                                                            4840
```

summarize weight, detail

		Weig	ht (lbs.)				
	Percentiles	Small	est				
1%	1760	1	760				
5%	1830	1	800				
10%	2020	1	800	0bs	74		
25%	2240	1	830	Sum of wgt.	74		
50%	3190			Mean	3019.459		
		Larg	est	Std. dev.	777.1936		
75%	3600	4	290				
90%	4060		330	Variance	604029.8		
95%	4290	4	720	Skewness	.1481164		
99%	4840	4:	840	Kurtosis	2.118403		
	oreign = Domesti						
	Variable	0bs		Std. dev.	Min	Max	
	price				3291	15906	
	oreign = Foreign						. = =
	Variable +					Max	
	price	22	6384.682	2621.915		12990	
f	oreign = .						
7	Variable	0bs	Mean	Std. dev.	Min	Max	
	price	0					

4 Tableaux de statistiques

4.1 Tableaux univariés

tabulate	foreign,	sort
----------	----------	------

Car origin	Freq.	Percent	Cum.
Domestic Foreign	52 22	70.27 29.73	70.27
Total	74	100.00	

tabulate foreign, summarize(price)

	Summary of Price					
Car origin	Mean	Std. dev.	Freq.			
+						
Domestic	6,072.423	3,097.104	52			
Foreign	6,384.682	2,621.915	22			
+						
Total	6,165.257	2,949.496	74			

tabstat price weight mpg, by(foreign) statistics(mean)

Summary statistics: Mean

Group variable: foreign (Car origin)

foreign	-	weight	mpg
Domestic Foreign	6072.423 6384.682	3317.115	19.82692 24.77273
•	6165.257		21.2973

tab1 foreign rep78 repeted

-> tabulation of foreign

Car origin	Freq.	Percent	Cum.
Domestic Foreign	52 22	70.27 29.73	70.27
Total	74	100.00	

-> tabulation of rep78

Repair	п	T	a
record 1978	Freq.	Percent	Cum.
1	2	2.90	2.90
2	8	11.59	14.49
3	30	43.48	57.97
4	18	26.09	84.06
5	11	15.94	100.00
+			
Total	69	100.00	

-> tabulation of repeted

Cum.	Percent	Freq.	repeted
86.67	86.67	65	0
98.67	12.00	9	1
100.00	1.33	1	2
			+
	100.00	75	Total

4.2 Tableaux bivariés

tab2 rep78 repeted foreign

-> tabulation of rep78 by repeted

Repair record		repeted		
1978	<u> </u>	0	1	Total
1	-+ 	2	0	2
2		8	0	8
3		28	2	30
4		14	4	18
5	j	8	3	11
	+		+	
Total	1	60	9	69

-> tabulation of rep78 by foreign

			Repair
	rigin	Car o	record
Total	Foreign	Domestic	1978
	+		+
2	0	2	1
8	0	8	2
30	3	27	3

1	9	9	4
	9	2	5
	21		+ Total

-> tabulation of repeted by foreign

	ı	Car c	rigin		
repeted		Domestic	Foreign		Total
	+-			+-	
0	ı	52	13	I	65
1	ı	0	9	I	9
	-+-			+-	
Total	I	52	22	1	74

tabulate rep78 foreign, row nofreq

Repair					
record		Car o	rigin		
1978		Domestic	Foreign		Total
	-+-			-+-	
1		100.00	0.00		100.00
2		100.00	0.00		100.00
3		90.00	10.00		100.00
4		50.00	50.00		100.00
5		18.18	81.82		100.00
	-+-			+-	
Total		69.57	30.43		100.00

tabulate rep78 foreign, summarize(price) means

Means of Price

Repair record		Car or	rigin		
1978	 -+-		Foreign		
1	1	4,564.5	•		4,564.5
2		5,967.625	•		5,967.625
3		6,607.074	4,828.667		6,429.233
4		5,881.556	6,261.444		6,071.5
5	1	4,204.5	6,292.667		5,913
Total	-+- 		6,070.143	٠	

4.3 Tableaux de statistiques

mean price weigh	t, over(forei	gn)		
Mean estimation			Number	of obs = 74
	Mean	Std. err.	[95% conf	. interval]
c.price@foreign	.+ 			
		429.4911	5216.449	6928.398
Foreign	6384.682	558.9942	5270.608	7498.756
c.weight@foreign				
Domestic	3317.115	96.4296	3124.931	3509.299
Foreign	2315.909	92.31665	2131.922	2499.896
total price weig	ht, over(fore	eign)		
Total estimation			Number	of obs = 74
		Std. err.		
c.price@foreign	•			
Domestic	315766	22333.54	271255.3	360276.7
Foreign	140463	12297.87	115953.4	164972.6
c.weight@foreign				
		5014.339	162496.4	182483.6
Foreign	50950	2030.966	46902.29	54997.71
proportion repet	ed, over(fore	eign)		
Proportion estima	ation		Number o	f obs = 74
 I			Log	 it
į	-	Std. err.	[95% conf.	
+ repeted@foreign				
O Domestic	1	0		_
•		.1048236	.3783643	774159
1 Domestic		(no observati		
		.1048236		.6216357
2 Domestic		(no observati		
•				

ratio ppoids: price/weight, over(foreign)

Ratio estimation Number of obs = 74

ppoids: price/weight

	 +	Ratio	Linearized std. err.	[95% conf.	interval]
c.ppoids@foreign Domestic Foreign	 	1.830634 2.756879	.1016378 .1528363	1.62807 2.452277	2.033198 3.061482

4.4 Tableaux généralisés

table (rep78)(foreign), statistic(mean price) statistic(median weight)

		1	Car origin	
		Domestic	Foreign	Total
		+		
Repair record	1 1978	I		
1				
Mean		1		
Price		4564.5		4564.5
Median				
Weight	(lbs.)	3100		3100
2				
Mean		1		
Price		5967.625		5967.625
Median				
Weight	(lbs.)	3465		3465
3		1		
Mean				
Price		6607.074	4828.667	6429.233
Median				
Weight	(lbs.)	3350	2070	3305
4		1		
Mean		1		
Price		5881.556	6261.444	6071.5
Median				
Weight	(lbs.)	3700	2160	2615
5				
Mean		1		

Price	1	4204.5	6292.667	5913
Median	(15-)	1060	2240	2222
Weight	(lbs.)	1960	2240	2200
Mean	l			
Price	į	6179.25	6070.143	6146.043
Median				
Weight	(lbs.)	3370	2160	3200

5 Création et modification de variables

6 Création de variables

```
generate cout = rep78 * 12500

(6 missing values generated)

generate loi = uniform()

generate marque = word(make, 1)

(1 missing value generated)

bysort rep78 foreign: egen vprice = mean(price)

(1 missing value generated)
```

6.1 Étiquettes (labels) de variables

```
label variable cout "Le coût de réparation"

ssc install elabel

checking elabel consistency and verifying not already installed...
all files already exist and are up to date.

elabel variable (loi marque)("Loi Uniforme" "La marque de la voiture")
```

6.2 Modifier une variable

```
replace loi = rnormal()
```

(75 real changes made)

```
ssc install ereplace
```

checking ereplace consistency and verifying not already installed... all files already exist and are up to date.

```
bysort rep78 foreign: ereplace vprice = total(price)
```

(74 real changes made)

6.3 Transformation en variables catégorielles

```
egen prix_cl = cut(price), at(3291, 5000, 10000, 15906) icodes
```

```
egen mpg_cl = cut(mpg), group(3)
```

(1 missing value generated)

```
generate weight_cl = autocode(weight, 4, 1760, 4840)
```

(1 missing value generated)

6.4 Étiquettes des valeurs

```
label define prix_cod 0 "Moins cher" 1 Abordable 2 Cher 3 "Très Cher"
```

```
label values prix_cl prix_cod
```

```
recode weight_cl (2530 = 1 "Légère")(3300 = 2 "Moins lourde") ///
(4070 = 3 "Lourde")(else = 4 "Très lourde"), generate(new_weight)
```

(75 differences between weight_cl and new_weight)

6.5 Transformation texte et numérique

```
decode foreign, generate(foreigntxt) maxlength(7)
```

```
label define fcode 1 Domesti 2 Foreign encode foreigntxt, generate(foreigncod) label(fcode)
```

```
tostring gear_ratio, generate(geartxt) force
```

geartxt generated as str11
geartxt was forced to string; some loss of information

```
destring geartxt, generate(gearnum) ignore("." "/") force
```

geartxt: character . removed; gearnum generated as double
(1 missing value generated)

6.6 Suppression de variables et observations

drop cout loi

keep make price mpg marque prix_cl foreign rep78

keep if !missing(price)

(1 observation deleted)

drop in 1/22

(22 observations deleted)

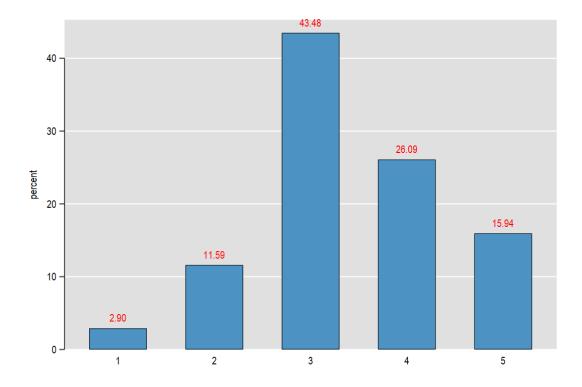
7 Illustration graphique

ssc install schemepack, replace

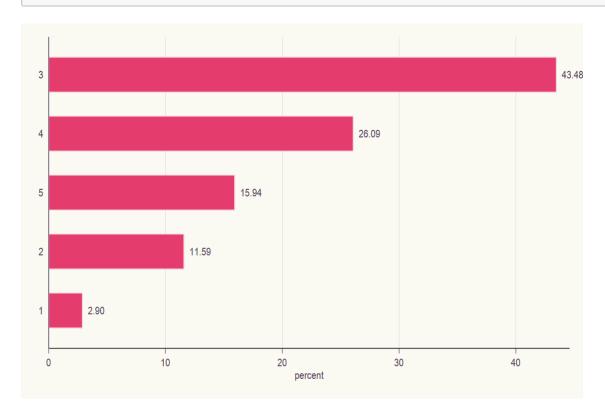
checking schemepack consistency and verifying not already installed... all files already exist and are up to date.

set scheme gg_tableau

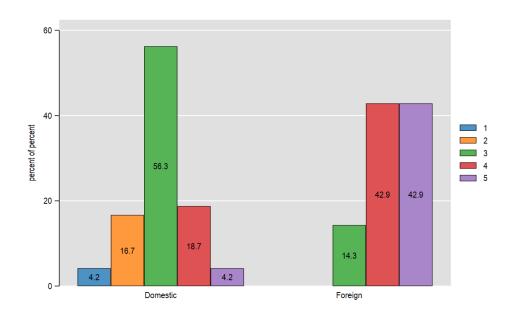
graph bar (percent), over(rep78) blabel(bar, format(%9.2f) color(red))



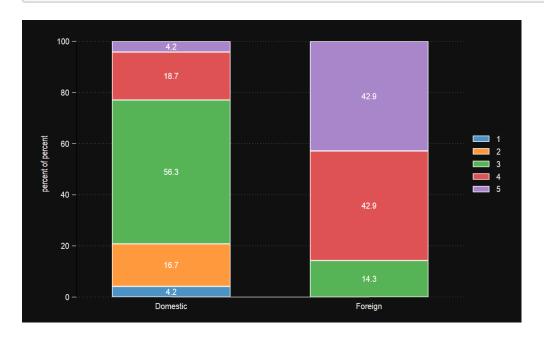
graph hbar (percent), over(rep78) blabel(bar, format(%9.2f)) scheme(swift_red)



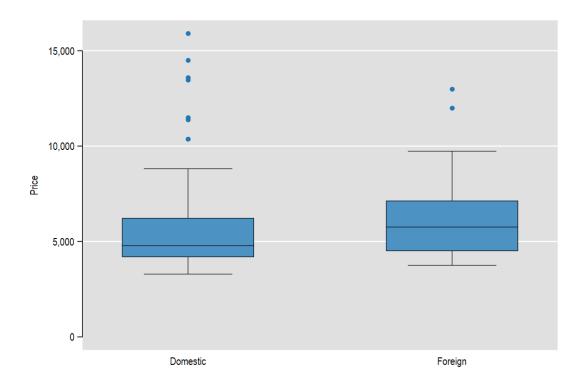
graph bar (percent), over(rep78) over(foreign) asyvars percentages ///
blabel(bar, position(center) format(%9.1f))



graph bar (percent), over(rep78) over(foreign) asyvars percentages stack ///
blabel(bar, position(center) format(%9.1f)) scheme(black_tableau)



graph box price, over(foreign)



graph pie, over(rep78) pie(_all, explode(10)) ///
plabel(_all percent, color(blue) format(%4.1f))

