Stata pour les débutants

Ibrahima TALL
Ingénieur statisticien économiste

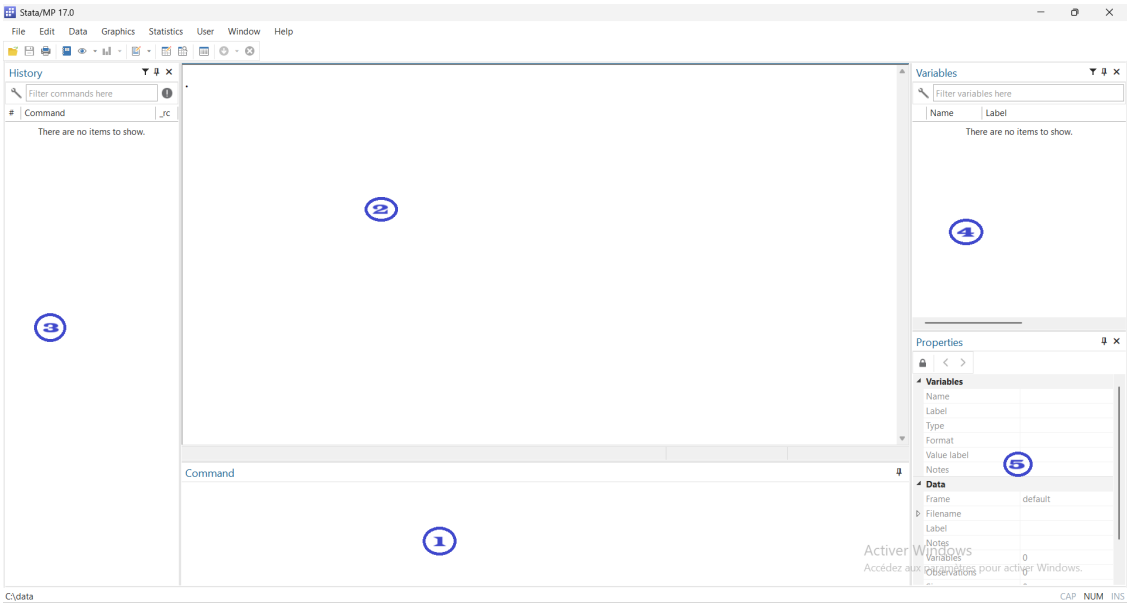
4 Avril 2025

---

1 Introduction

Stata est 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, c'est la version 18 qui en est la dernière. Dans cette présentation, c'est la version 17 qui est utilisée.

1.1 Fenêtres de l'interface



1. insertion des commandes validées par la touche entrée ;
2. sorties des résultats ;
3. historiques des actions ;
4. variables avec leurs étiquettes ;
5. propriétés de la base et des variables.

1.2 Commandes stata

```
ssc new
```

```
(contacting http://repec.org)
```

```
ssc whatshot, n(10)
```

```
Top 10 packages at SSC

       Feb 2025
  Rank   # hits    Package       Author(s)
  ------------------------------------------------------------------------
     1  87128.6    estout        Ben Jann
     2  83967.2    outreg2       Roy Wada
     3  68888.4    reghdfe       Sergio Correia
     4  67323.3    winsor2       Yujun Lian
     5  64700.7    asdoc         Attaullah Shah
     6  44584.0    ftools        Sergio Correia
     7  33506.2    ivreg210      Christopher F Baum, Mark E Schaffer,
                                    Steven Stillman
     8  31413.5    ivreg2        Christopher F Baum, Steven Stillman,
                                    Mark E Schaffer
     9  30169.0    sum2docx      Chuntao Li, Yuan Xue
    10  29850.2    ivreg29       Steven Stillman, Mark E Schaffer,
                                    Christopher F Baum
  ------------------------------------------------------------------------
  (Click on package name for description)
```

```
ssc hot, n(5)
```

```
Top 5 packages at SSC

       Feb 2025
  Rank   # hits    Package       Author(s)
  ------------------------------------------------------------------------
     1  87128.6    estout        Ben Jann
     2  83967.2    outreg2       Roy Wada
     3  68888.4    reghdfe       Sergio Correia
     4  67323.3    winsor2       Yujun Lian
     5  64700.7    asdoc         Attaullah Shah
  ------------------------------------------------------------------------
  (Click on package name for description)
```

```
ssc describe d
```

```
------------------------------------------------------------------------------
http://fmwww.bc.edu/repec/bocode/d/
(no title)
------------------------------------------------------------------------------

PACKAGES you could -net describe-:
    d3network         module to create network visualizations using D3.js to
                        view in browser
    dummies           function to create families of dummy variables
    dummies2          module to create indicator variables from categorical
                        variable and vice versa
    dummieslab        module to convert categorical variable to dummy
                        variables using label names
    duncan            module to calculate dissimilarity index
    dups              module to identify and optionally remove duplicate
                        observations
------------------------------------------------------------------------------
(type ssc describe pkgname for more information on pkgname)
```

```
ssc describe dummies
```

```
------------------------------------------------------------------------------
package dummies from http://fmwww.bc.edu/repec/bocode/d
------------------------------------------------------------------------------

TITLE
      'DUMMIES': function to create families of dummy variables

DESCRIPTION/AUTHOR(S)

      dummies creates families of dummy variables for each variable in
      varlist. Applied to varname, it produces varname1, varname2, etc.
      Applied to longname (a name 8 characters long), it produces
      longnam1, longnam2, etc. This is version 1.1.1 of  the software.

      Author: Nicholas J. Cox, University of Durham
      Support: email N.J.Cox@durham.ac.uk

      Distribution-Date: 19980923

INSTALLATION FILES                                (type net install dummies)
      dummies.ado
      dummies.hlp
------------------------------------------------------------------------------
(type ssc install dummies to install)
```

```
search dummies
```

```
ssc install dummies
```

```
checking dummies consistency and verifying not already installed...
installing into C:\Users\IBRAHIMA TALL\ado\plus\...
installation complete.
```

```
which dummies
```

```
C:\Users\IBRAHIMA TALL\ado\plus\d\dummies.ado
*! NJC 1.1.1 20 Oct 1998
```

```
net install dups
```

```
checking dups consistency and verifying not already installed...
installing into C:\Users\IBRAHIMA TALL\ado\plus\...
installation complete.
```

```
ssc uninstall dups
```

```
package dups from http://fmwww.bc.edu/repec/bocode/d
      'DUPS': module to identify and optionally remove duplicate observations

(package uninstalled)
```

```
help dummies
```

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

```
cd ..
```

```
C:\Users\IBRAHIMA TALL\Documents
```

```
dir
```

```
  <dir>   3/26/25 22:22  .
  <dir>   2/11/25 10:34  ..
  <dir>   3/23/25 10:39  .ipynb_checkpoints
   0.0k   3/09/25 16:12  .Rhistory
   6.2k   3/26/25 15:27  Calcul.xlsx
  <dir>   2/07/25  1:12  Ma musique
  <dir>   2/07/25  1:12  Mes images
  <dir>   2/07/25  1:12  Mes vidéos
  <dir>  10/01/24 13:42  Modèles Office personnalisés
   0.0k   3/26/25 15:26  Monfichier.txt
  10.3k  10/05/24  1:41  MPN.png
```

```
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 // commande 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 EXPOSES
```

```
mkdir mondoc
```

```
rmdir mondoc
```

```
cd ./EXPOSES
```

```
C:\Users\IBRAHIMA TALL\Documents\EXPOSES
```

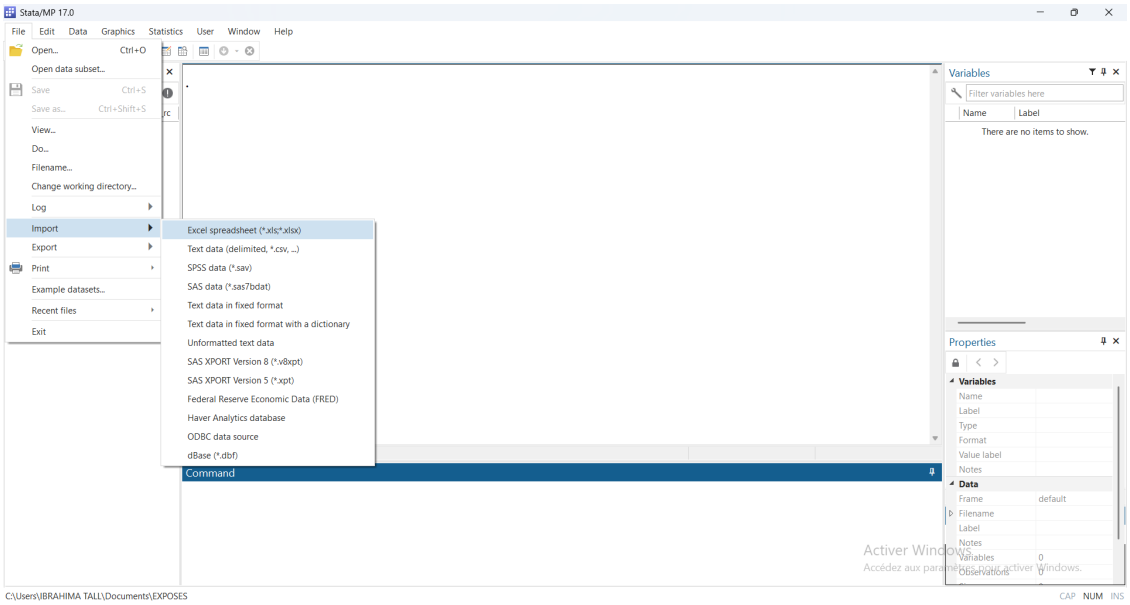
```
which display
```

```
built-in command:  display
```

2 Usage des bases de données

L'importer ou d'exporter des données peut se faire via l'interface ou directement avec les commandes. L'interface est plus commode 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:

- stata (.dta): en mémoire: use, en système: sysuse dir et sysuse, en ligne: webuse ;
- type excel: import *excel*;
- type texte: import *delimited* ;
- type SPSS (.sav): import *spss*.

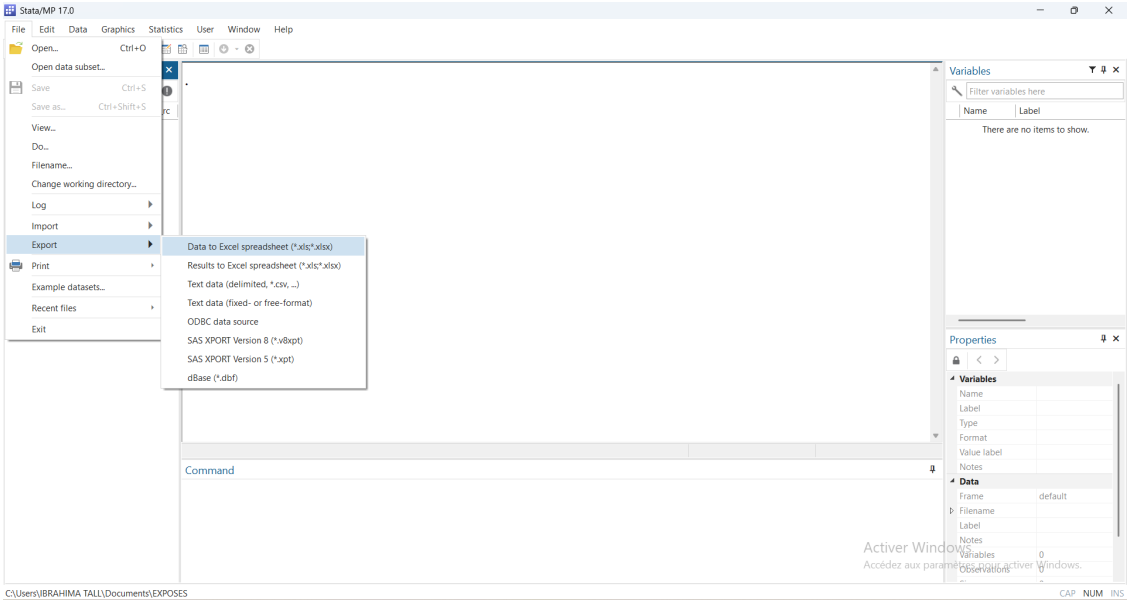
```
sysuse dir
```

```
  auto.dta        census.dta      network1.dta    surface.dta
  auto16.dta      citytemp.dta    network1a.dta   tsline1.dta
  auto2.dta       citytemp4.dta   nlsw88.dta      tsline2.dta
  autornd.dta     educ99gdp.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 import par export. Pour les données de type stata, l'exportation correspond à une sauvegarde avec la commande save.

```
save mabase, replace
```

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

2.3 Combinaison de bases de données

```
isid make
```

```
merge 1:1 make using mabase.dta
```

```
(label origin already defined)

    Result                      Number of obs
    -----------------------------------------
    Not matched                             0
    Matched                                74  (_merge==3)
    -----------------------------------------
```

```
append using mabase.dta, generate(linked)
```

```
(label origin already defined)
```

3 Exploitation de la base

```
use mabase, clear
```

```
(1978 automobile data)
```

3.1 Noms de variables

```
describe, simple
```

```
make  mpg  headroom  weight  turn  gear_ratio
price rep78 trunk length displacement foreign
```

```
rename price cout
rename * v#, addnumber
rename * v#, addnumber(20)
rename v# v#, renumber(0) sort
rename (v1 v2 v3)(Marque cout kilometrage)
rename v?, upper
rename *, lower
```

3.2 Observation 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      Storage   Display    Value
    name         type    format    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 foreign
```

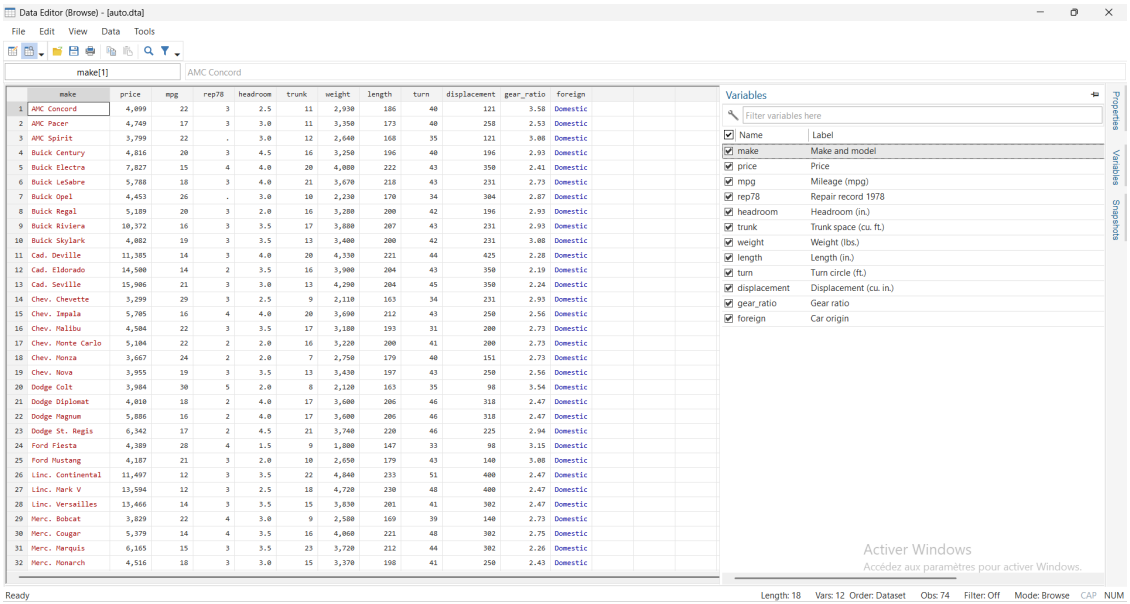
```
-------------------------------------------------------------------------------
price                                                                   Price
-------------------------------------------------------------------------------

                  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

-------------------------------------------------------------------------------
foreign                                                           Car origin
-------------------------------------------------------------------------------

                  Type: Numeric (byte)
                 Label: origin

                 Range: [0,1]                         Units: 1
         Unique values: 2                          Missing .: 0/74

            Tabulation: Freq.   Numeric  Label
                          52         0  Domestic
                          22         1  Foreign
```

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

```
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
```



3.3 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
```

3.4 Rangement de la base

```
sort make
```

```
gsort foreign -price
```

```
order turn foreign, after(make)
```

```
order foreign, last
```

3.5 Ajout d'observations

```
set obs 75
```

```
Number of observations (_N) was 74, now 75.
```

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

```
(2 observations added)
```

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

```
(9 observations created)
```

3.6 Doublons sur les observations

```
duplicates report make
```

```
Duplicates in terms of make

-------------------------------------
   Copies | Observations       Surplus
----------+--------------------------
        1 |           65             0
        2 |           18             9
        3 |            3             2
-------------------------------------
```

```
duplicates list make
```

```
Duplicates in terms of make

  +------------------------------+
  | Group   Obs   make           |
  |------------------------------|
  |     1    20                  |
  |     1    21                  |
  |     1    77                  |
  |     2    70   Datsun 210     |
  |     2    82   Datsun 210     |
  |------------------------------|
  |     3    68   Datsun 510     |
  |     3    80   Datsun 510     |
  |     4    72   Fiat Strada    |
  |     4    84   Fiat Strada    |
  |     5    71   Honda Civic    |
  |------------------------------|
  |     5    83   Honda Civic    |
  |     6    73   Mazda GLC      |
  |     6    85   Mazda GLC      |
  |     7    74   Renault Le Car |
  |     7    86   Renault Le Car |
  |------------------------------|
  |     8    66   Toyota Corona  |
  |     8    78   Toyota Corona  |
  |     9    67   VW Diesel      |
  |     9    79   VW Diesel      |
  |    10    69   VW Rabbit      |
  |------------------------------|
  |    10    81   VW Rabbit      |
  +------------------------------+
```

```
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)
```

3.7 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
```

3.8 Statistiques usuelles

```
count if price <= 5000
```

```
  37
```

```
by foreign, sort: count if price <= 5000
```

```
-------------------------------------------------------------------------------
-> foreign = Domestic
  29
-------------------------------------------------------------------------------
-> foreign = Foreign
  8
-------------------------------------------------------------------------------
-> foreign = .
  0
```

```
inspect price
```

```
price:  Price                             Number of observations
------------                       ----------------------------------------
                                          Total      Integers   Nonintegers
|  #                        Negative            -             -             -
|  #                        Zero                -             -             -
|  #                        Positive           74            74             -
|  #                                   -----------   -----------   -----------
|  #                        Total              74            74             -
|  #   #   .   .   .        Missing             1
+----------------------              -----------
3291              15906                       75
  (74 unique values)
```

```
bysort foreign: summarize price
```

```
-------------------------------------------------------------------------------
-> foreign = Domestic

    Variable |        Obs        Mean    Std. dev.       Min        Max
-------------+---------------------------------------------------------
       price |         52    6072.423    3097.104       3291      15906

-------------------------------------------------------------------------------
-> foreign = Foreign

    Variable |        Obs        Mean    Std. dev.       Min        Max
-------------+---------------------------------------------------------
       price |         22    6384.682    2621.915       3748      12990

-------------------------------------------------------------------------------
-> foreign = .

    Variable |        Obs        Mean    Std. dev.       Min        Max
-------------+---------------------------------------------------------
       price |          0
```

```
statsby, basepop(!foreign) by(rep78): summarize price
```

4 Tableaux de statistiques

4.1 Tableaux univariés

```
tabulate foreign, sort
```

```
 Car origin |      Freq.     Percent        Cum.
------------+-----------------------------------
   Domestic |         52       70.27       70.27
    Foreign |         22       29.73      100.00
------------+-----------------------------------
      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 |     price    weight       mpg
---------+------------------------------
Domestic |  6072.423  3317.115  19.82692
 Foreign |  6384.682  2315.909  24.77273
---------+------------------------------
   Total |  6165.257  3019.459   21.2973
----------------------------------------
```

```
tab1 foreign rep78 repeted
```

```
-> tabulation of foreign

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

-> tabulation of rep78

     Repair |
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

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

4.2 Tableaux bi-variés

```
tab2 rep78 repeted foreign
```

```
-> tabulation of rep78 by repeted

    Repair |
    record |        repeted
      1978 |         0          1 |     Total
-----------+----------------------+----------
         1 |         2          0 |         2
         2 |         8          0 |         8
         3 |        28          2 |        30
         4 |        14          4 |        18
         5 |         8          3 |        11
-----------+----------------------+----------
     Total |        60          9 |        69

-> tabulation of rep78 by foreign

    Repair |
    record |      Car origin
      1978 |  Domestic    Foreign |     Total
-----------+----------------------+----------
         1 |         2          0 |         2
         2 |         8          0 |         8
         3 |        27          3 |        30
         4 |         9          9 |        18
         5 |         2          9 |        11
-----------+----------------------+----------
     Total |        48         21 |        69

-> tabulation of repeted by foreign

           |      Car origin
   repeted |  Domestic    Foreign |     Total
-----------+----------------------+----------
         0 |        52         13 |        65
         1 |         0          9 |         9
-----------+----------------------+----------
     Total |        52         22 |        74
```

```
tabulate rep78 foreign, row nofreq
```

```
    Repair |
    record |      Car origin
      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 origin
      1978 |  Domestic    Foreign |     Total
-----------+----------------------+----------
         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 |   4,204.5  6,292.667 |     5,913
-----------+----------------------+----------
     Total |  6,179.25  6,070.143 | 6,146.043
```

4.3 Tableaux de statistiques

```
mean price weight, over(foreign)
```

```
Mean estimation                                 Number of obs = 74

------------------------------------------------------------------------
                 |       Mean   Std. err.     [95% conf. interval]
-----------------+------------------------------------------------------
 c.price@foreign |
       Domestic  |   6072.423   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 weight, over(foreign)
```

```
Total estimation                                  Number of obs = 74

------------------------------------------------------------------------
                  |      Total   Std. err.     [95% conf. interval]
------------------+-----------------------------------------------------
 c.price@foreign  |
       Domestic   |     315766   22333.54      271255.3    360276.7
        Foreign   |     140463   12297.87      115953.4    164972.6
                  |
c.weight@foreign  |
       Domestic   |     172490   5014.339      162496.4    182483.6
        Foreign   |      50950   2030.966      46902.29    54997.71
------------------------------------------------------------------------
```

```
proportion repeted, over(foreign)
```

```
Proportion estimation                             Number of obs = 74

------------------------------------------------------------------------
                  |                                      Logit
                  | Proportion   Std. err.     [95% conf. interval]
------------------+-----------------------------------------------------
repeted@foreign   |
    0 Domestic    |          1          0              .           .
     0 Foreign    |   .5909091   .1048236      .3783643     .774159
    1 Domestic    |          0  (no observations)
     1 Foreign    |   .4090909   .1048236       .225841    .6216357
    2 Domestic    |          0  (no observations)
     2 Foreign    |          0  (no observations)
------------------------------------------------------------------------
```

```
ratio ppoids: price/weight, over(foreign)
```

```
Ratio estimation                                  Number of obs = 74

       ppoids: price/weight

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

4.4 Tableaux généralisés

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

```
----------------------------------------------------
                    |            Car origin
                    |  Domestic    Foreign      Total
--------------------+-------------------------------
Repair record 1978  |
  1                 |
    Mean            |
      Price         |    4564.5                4564.5
    Median          |
      Weight (lbs.) |      3100                  3100
  2                 |
    Mean            |
      Price         |  5967.625              5967.625
    Median          |
      Weight (lbs.) |      3465                  3465
  3                 |
    Mean            |
      Price         |  6607.074   4828.667   6429.233
    Median          |
      Weight (lbs.) |      3350       2070       3305
  4                 |
    Mean            |
      Price         |  5881.556   6261.444     6071.5
    Median          |
      Weight (lbs.) |      3700       2160       2615
  5                 |
    Mean            |
      Price         |    4204.5   6292.667       5913
    Median          |
      Weight (lbs.) |      1960       2240       2200
  Total             |
    Mean            |
      Price         |   6179.25   6070.143   6146.043
    Median          |
      Weight (lbs.) |      3370       2160       3200
----------------------------------------------------
```

5 Création et modification de variables

5.1 Création de variables

```
generate cout = rep78 * 12500
```

```
(6 missing values generated)
```

```
generate loi1 = runiform()
```

```
generate loi2 = runiform()
```

```
compare loi1 loi2
```

```
                                        ---------- Difference ----------
                            Count       Minimum      Average     Maximum
-------------------------------------------------------------------------
loi1<loi2                      35     -.8829869    -.2928788   -.0008371
loi1>loi2                      39       .046526     .3743599    .7960947
                          ----------
Jointly defined                74     -.8829869     .0587741    .7960947
                          ----------
Total                          74
```

```
set seed 123456
generate loi1u = runiform()
```

```
set seed 123456
generate loi2u = runiform()
```

```
compare loi2u loi1u
```

```
                                        ---------- Difference ----------
                            Count       Minimum      Average     Maximum
-------------------------------------------------------------------------
loi2u=loi1u                    74
                          ----------
Jointly defined                74             0            0           0
                          ----------
Total                          74
```

```
generate marque = word(make, 1)
```

```
(1 missing value generated)
```

```
bysort rep78 foreign: egen vprice = mean(price)
```

```
(1 missing value generated)
```

5.2 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")
```

5.3 Modification 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)
```

5.4 Transformation en variables catégorielles

```
generate cpoids = cond(mpg <= 20, 1, 2)
```

```
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)
```

5.5 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)
```

5.6 Transformation en texte et en 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)
```

5.7 Suppression de variables et d'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)
```

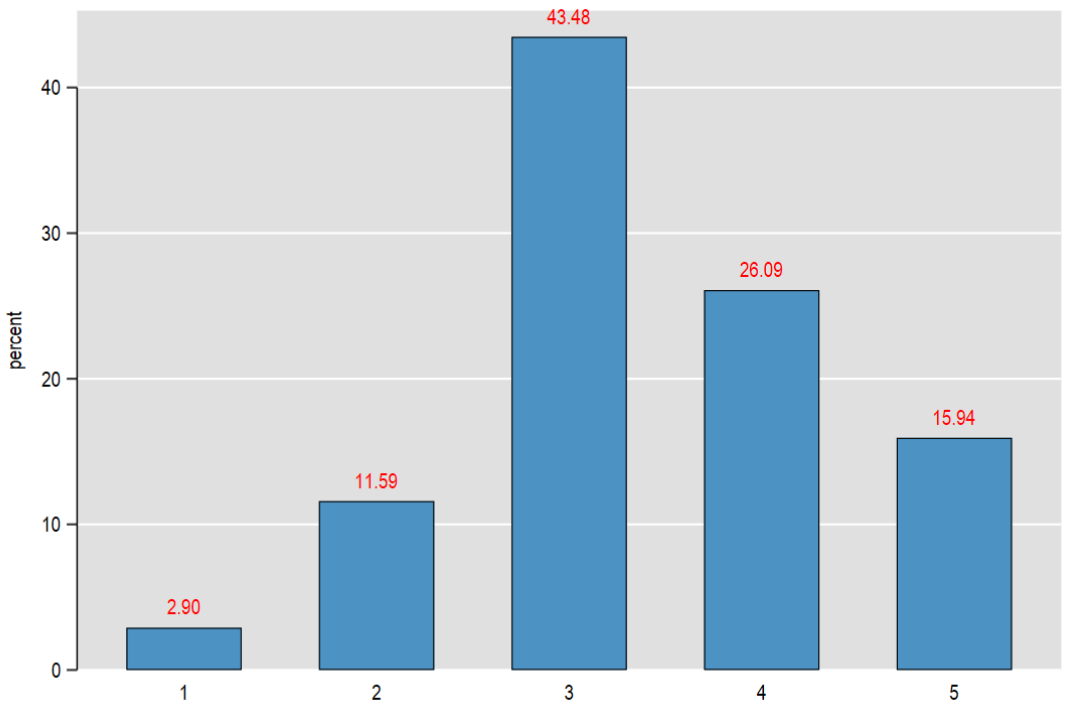
6 Illustration graphique

```
ssc install schemepack, replace
```

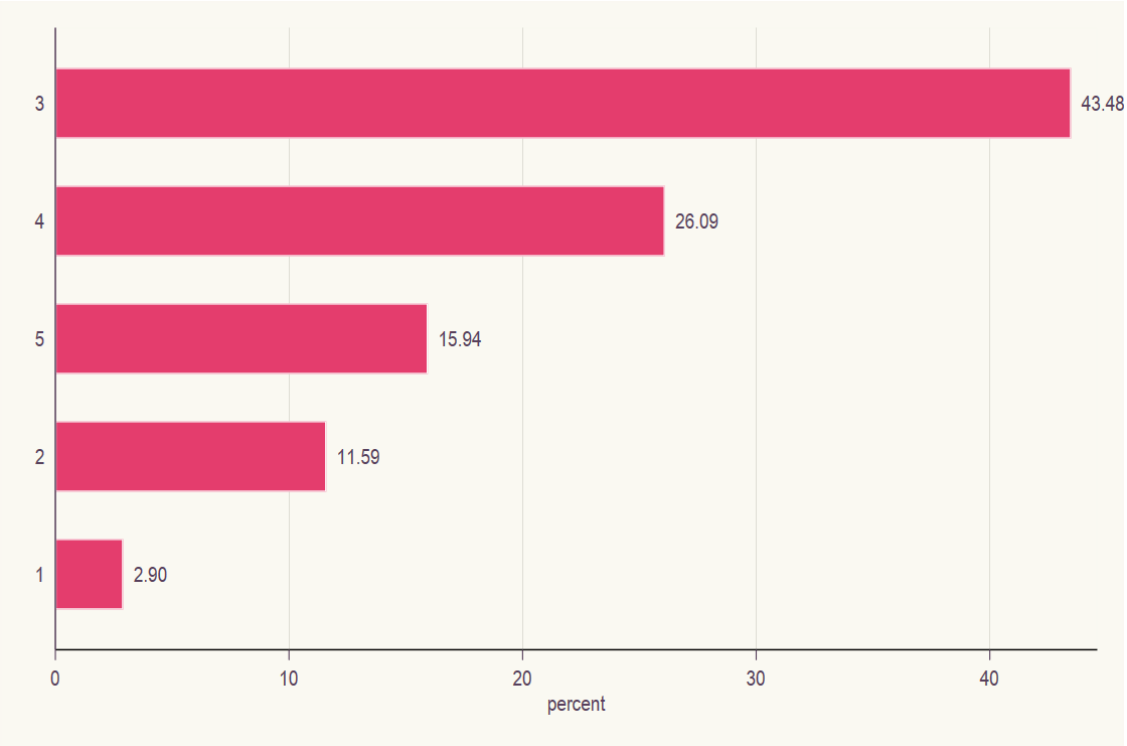
```
checking schemepack consistency and verifying not already installed...
all files already exist and are up to date.
```

```
set scheme gg_tableau
```

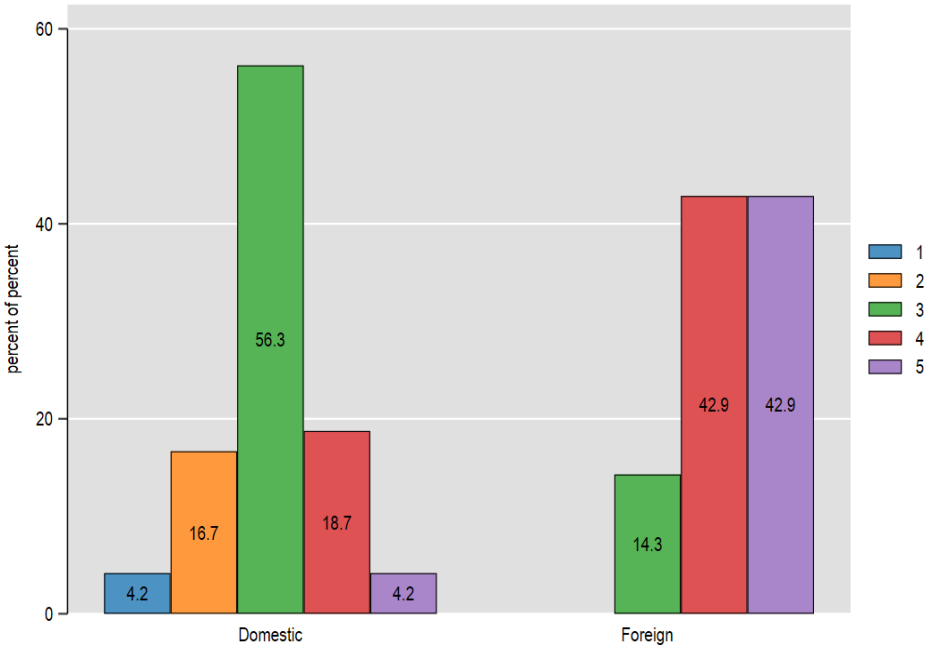
```
graph bar (percent), over(rep78) blabel(bar, format(%9.2f) color(red))
```



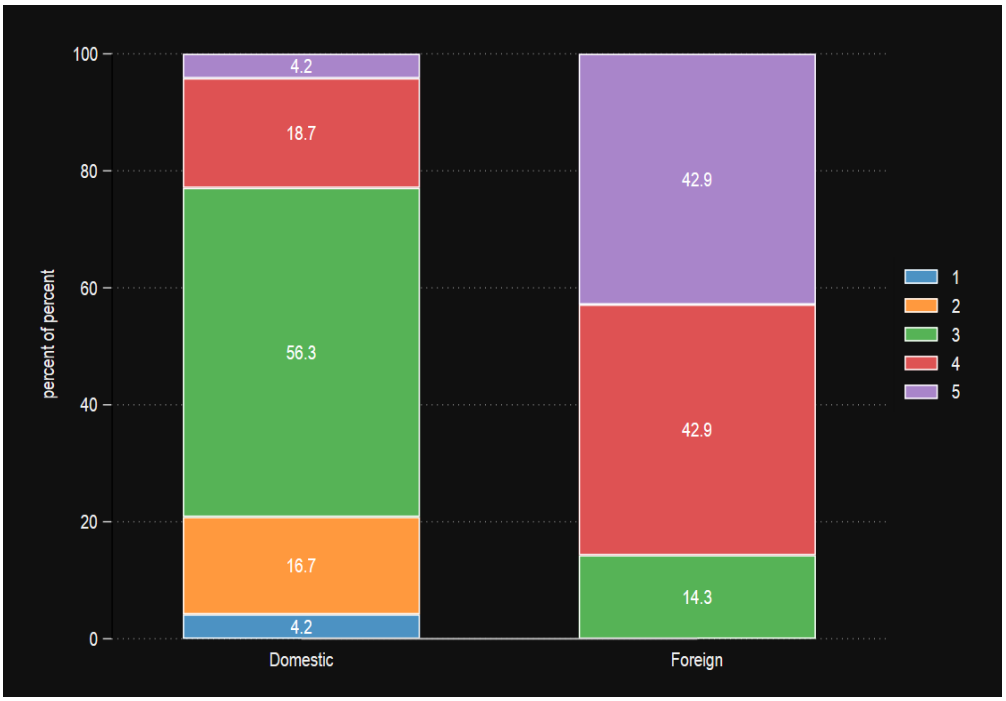

```
graph hbar (percent), over(rep78) blabel(bar, format(%9.2f)) scheme(swift_red)
```



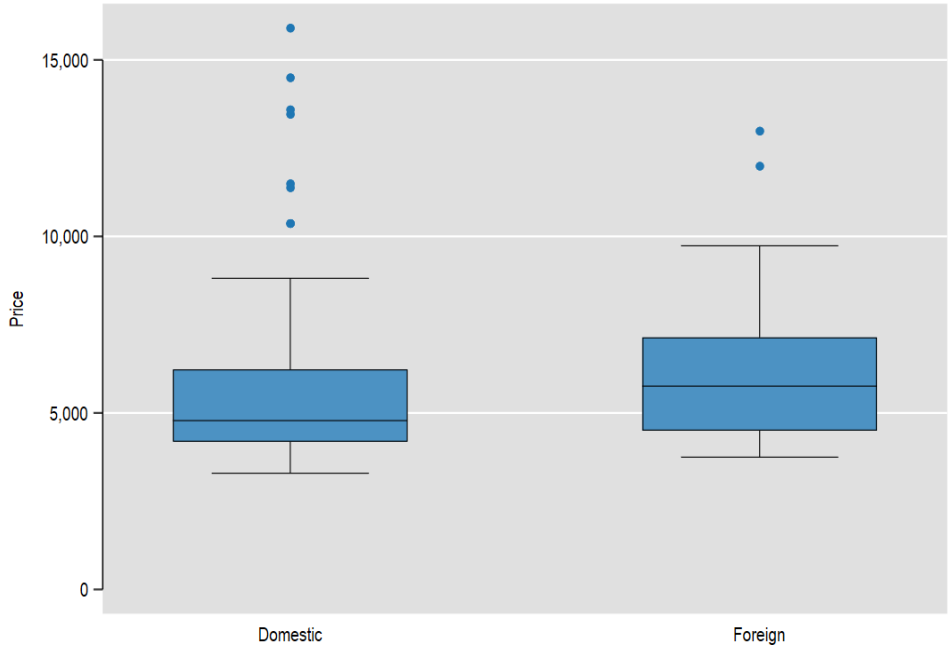
```
graph bar (percent), over(rep78) over(foreign) asyvars percentages ///
 blabel(bar, position(center) format(%9.1f))
```




```
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))
```