

## COURSE

# Quản lý & Biên tập số liệu với Stata

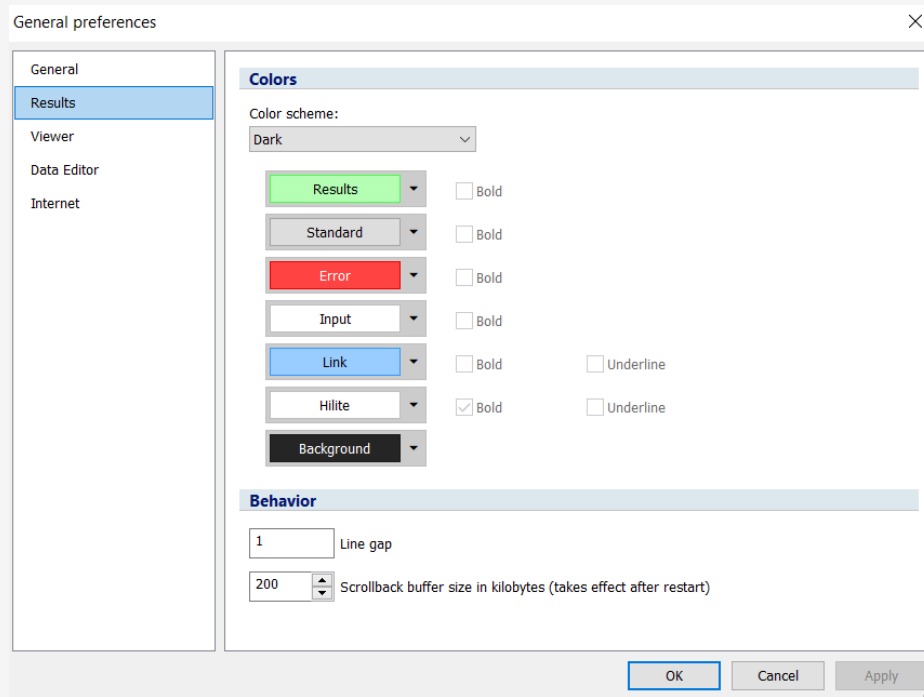
---

Lớp phân tích thống kê cơ bản

Khương Quỳnh Long  
Hà Nội, 06-08/06/2020

# Giao diện Stata

- **Chỉnh màn hình màu**
  - Edit → Preferences  
→ General Preferences
  - Trong mục Color scheme, chọn giao diện phù hợp
  - Tắt Stata và khởi động lại Stata



Stata/SE 16.0 - D:\Dropbox\Long\Projects\Courses\Basic\Thông kê mô tả\sl7\_huyetap.dta

File Edit Data Graphics Statistics User Window Help

History

Filter commands here

# Command \_rc

1 use "D:\Dropbox\Long\Proj...

Statistics/Data Analysis 16.0 Copyright 1985-2019 StataCorp LLC  
StataCorp  
4905 Lakeway Drive  
College Station, Texas 77845 USA  
800-STATA-PC http://www.stata.com  
979-696-4600 stata@stata.com  
979-696-4601 (fax)

Special Edition

Single-user Stata license expires 19 Mar 2020:  
Serial number: 401609201780  
Licensed to: Khuong Quynh Long  
Hanoi University of Public Health

Notes:

1. Unicode is supported; see help unicode\_advice.
2. Maximum number of variables is set to 5000; see help set\_maxvar.

. use "D:\Dropbox\Long\Projects\Courses\Basic\Thông kê mô tả\sl7\_huyetap.dta"

Command

Cửa sổ kết quả

Đánh câu lệnh

Variables

Filter variables here

| Name        | Label                 |
|-------------|-----------------------|
| ma          | ma ca nhan            |
| tuoi        | tuoi                  |
| gioitinh    | gioi tinh             |
| cannang     | can nang              |
| caotb       | chieu cao tb          |
| suckhoe     | nhân xét về sức khỏe  |
| ttthainghen | tình trạng thai nghen |
| caoha       | cao huyết áp          |
| chieuca1    | chieu cao lan 1       |
| chieuca2    | chieu cao lan 2       |
| hatoida1    | huyết áp tối đa 1     |
| hatoida2    | huyết áp tối đa 2     |
| hatoithieu1 | huyết áp tối thiểu 1  |
| hatoithieu2 | huyết áp tối thiểu 2  |
| nhomtuoi    | nhom tuoi             |
| hatdtb      | huyết áp tối đa tb    |
| hatttb      | huyết áp tối thiểu tb |

Properties

Variables

| Name        | Label |
|-------------|-------|
| Label       |       |
| Type        |       |
| Format      |       |
| Value label |       |
| Notes       |       |

Data

| Frame    | Label           |
|----------|-----------------|
| Filename | sl7_huyetap.dta |
| Label    |                 |
| Notes    |                 |

Thông tin biến

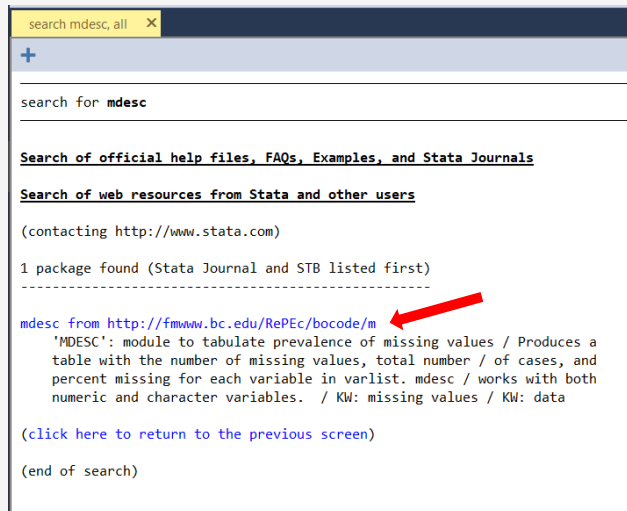
# Trợ giúp

---

- `help {tên câu lệnh}`
- Ví dụ: `help tab1`
- Tìm hiểu chức năng của câu lệnh sau
  - ✓ `tab, tab1, sum, tabstat, ci, hist`

# Cài thêm câu lệnh

- Cách 1: `ssc install {tên gói}`
- Cách 2: `findit {tên câu lệnh}`
- Ví dụ: `findit mdesc`



search mdesc, all

search for **mdesc**

Search of official help files, FAQs, Examples, and Stata Journals

Search of web resources from Stata and other users

(contacting <http://www.stata.com>)

1 package found (Stata Journal and STB listed first)

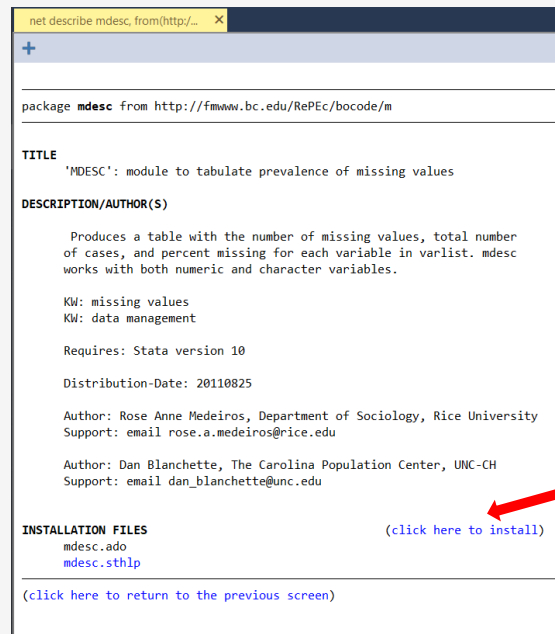
-----

[mdesc from http://fmwww.bc.edu/RePEc/bocode/m](http://fmwww.bc.edu/RePEc/bocode/m)

'MDESC': module to tabulate prevalence of missing values / Produces a table with the number of missing values, total number / of cases, and percent missing for each variable in varlist. mdesc / works with both numeric and character variables. / KW: missing values / KW: data

([click here to return to the previous screen](#))

(end of search)



net describe mdesc, from(<http://fmwww.bc.edu/RePEc/bocode/m>)

package **mdesc** from <http://fmwww.bc.edu/RePEc/bocode/m>

**TITLE**

'MDESC': module to tabulate prevalence of missing values

**DESCRIPTION/AUTHOR(S)**

Produces a table with the number of missing values, total number of cases, and percent missing for each variable in varlist. mdesc works with both numeric and character variables.

KW: missing values  
KW: data management

Requires: Stata version 10

Distribution-Date: 20110825

Author: Rose Anne Medeiros, Department of Sociology, Rice University  
Support: email [rose.a.medeiros@rice.edu](mailto:rose.a.medeiros@rice.edu)

Author: Dan Blanchette, The Carolina Population Center, UNC-CH  
Support: email [dan\\_blanchette@unc.edu](mailto:dan_blanchette@unc.edu)

**INSTALLATION FILES**

[mdesc.ado](#)  
[mdesc.sthlp](#)

([click here to install](#))

([click here to return to the previous screen](#))

# Nội dung

---

1. Mở và nhập số liệu
2. Xem cấu trúc số liệu
3. Tìm kiếm và chỉnh sửa số liệu
4. Sắp xếp số liệu
5. Lưu và xuất số liệu
6. Tạo biến mới
7. Đổi tên, xóa biến số
8. Tạo và gán nhãn
9. Thay thế giá trị số liệu
10. Phân nhóm cho biến số
11. Chuyển đổi dạng số liệu
12. Gộp bộ số liệu
13. Loop

# Mã hóa số liệu

```
. tab caoha
```

| cao huyet ap | Freq. | Percent | Cum.   |
|--------------|-------|---------|--------|
| khong cao ha | 319   | 80.35   | 80.35  |
| cao ha       | 78    | 19.65   | 100.00 |
| Total        | 397   | 100.00  |        |

```
. tab caoha, nol
```

| cao huyet ap | Freq. | Percent | Cum.   |
|--------------|-------|---------|--------|
| 0            | 319   | 80.35   | 80.35  |
| 1            | 78    | 19.65   | 100.00 |
| Total        | 397   | 100.00  |        |

- Mã hóa dữ liệu
- ✓ Tên biến số
- ✓ Nhãn biến số (label variable)
- ✓ Giá trị (value)
- ✓ Nhãn giá trị (label value)

# Mã hóa số liệu

---

- **Tên biến số**

- ✓ Không bắt đầu bằng số,
- ✓ Không bắt đầu bằng khoảng trắng hoặc ký tự đặc biệt
- ✓ Không dùng tiếng Việt có dấu
- ✓ Ngắn gọn

- **Nhãn biến số**

- ✓ Ngắn gọn
- ✓ Có thể dùng tiếng Việt có dấu
- ✓ Nên kèm theo đơn vị đo lường



# Mã hóa số liệu

---

- **Giá trị**

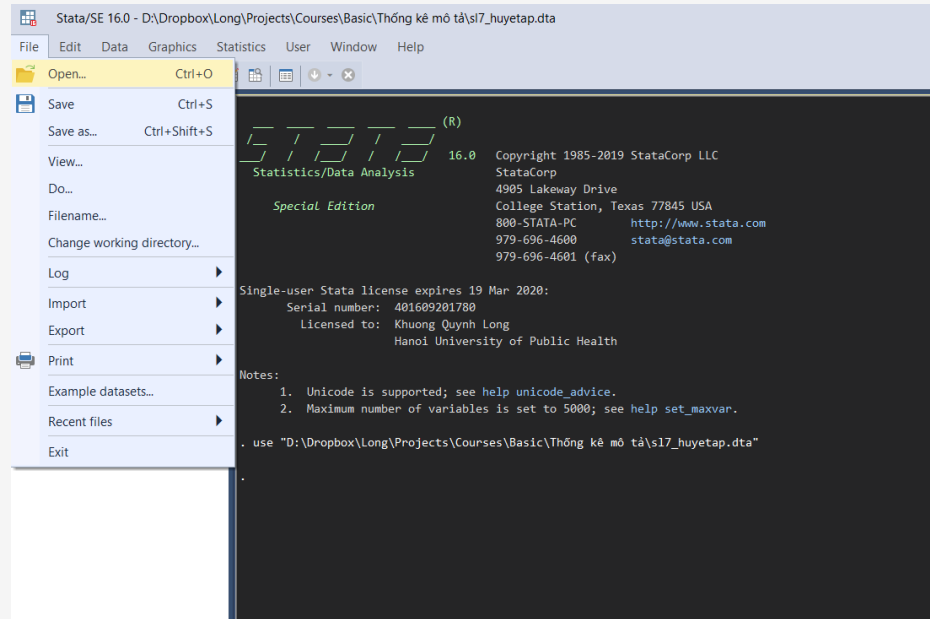
- ✓ Biến nhị giá: nên mã hóa 0/1
- ✓ Biến danh mục, thứ tự: mã hóa 1,2,3...
- ✓ Giá trị khuyết, từ chối trả lời: 9, 99, 999....

- **Nhãn giá trị**

- ✓ Ngắn gọn
- ✓ Có thể dùng tiếng Việt có dấu
- ✓ Không nên kèm theo đơn vị đo lường

# 1. Mở số liệu

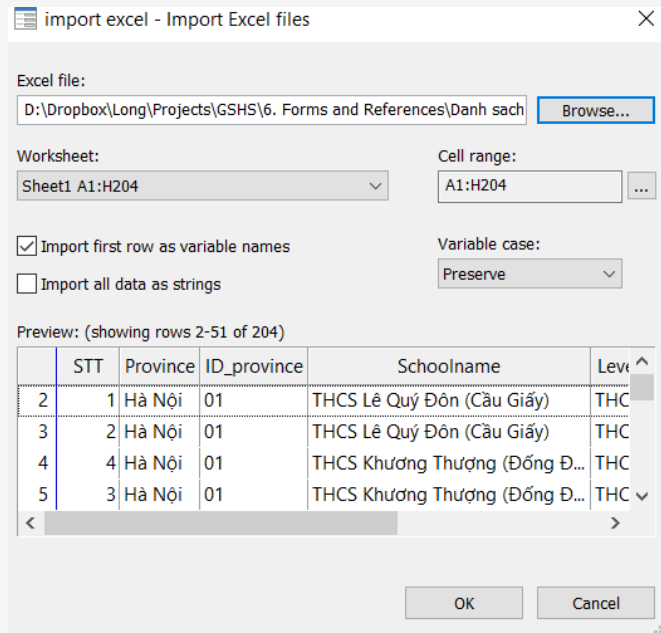
- File >> open >> chọn tập số liệu
- Tại 1 thời điểm, Stata chỉ mở được 1 bộ số liệu
- Cách dùng lệnh  
use {đường dẫn}/data.dta, clear



# 1. Mở số liệu

- Mở file excel

File >> import >> Excel spreadsheet(\*.xls; \*.xlsx)



import excel - Import Excel files

Excel file:  
D:\Dropbox\Long\Projects\GSHS\6. Forms and References\Danh sach Browse...

Worksheet:  
Sheet1 A1:H204

Cell range:  
A1:H204

☒ Import first row as variable names  
☐ Import all data as strings

Variable case:  
Preserve

Preview: (showing rows 2-51 of 204)

|   | STT | Province | ID_province | Schoolname                    | Level |
|---|-----|----------|-------------|-------------------------------|-------|
| 2 | 1   | Hà Nội   | 01          | THCS Lê Quý Đôn (Cầu Giấy)    | THC   |
| 3 | 2   | Hà Nội   | 01          | THCS Lê Quý Đôn (Cầu Giấy)    | THC   |
| 4 | 4   | Hà Nội   | 01          | THCS Khương Thượng (Đồng Đ... | THC   |
| 5 | 3   | Hà Nội   | 01          | THCS Khương Thượng (Đồng Đ... | THC   |

OK Cancel

## 2. Xem cấu trúc số liệu

---

***contain data from:*** nơi lưu trữ file số liệu

***obs:*** Số đối tượng trong nghiên cứu

***vars:*** Số biến số trong nghiên cứu

***Variable name:*** tên biến số trong số liệu

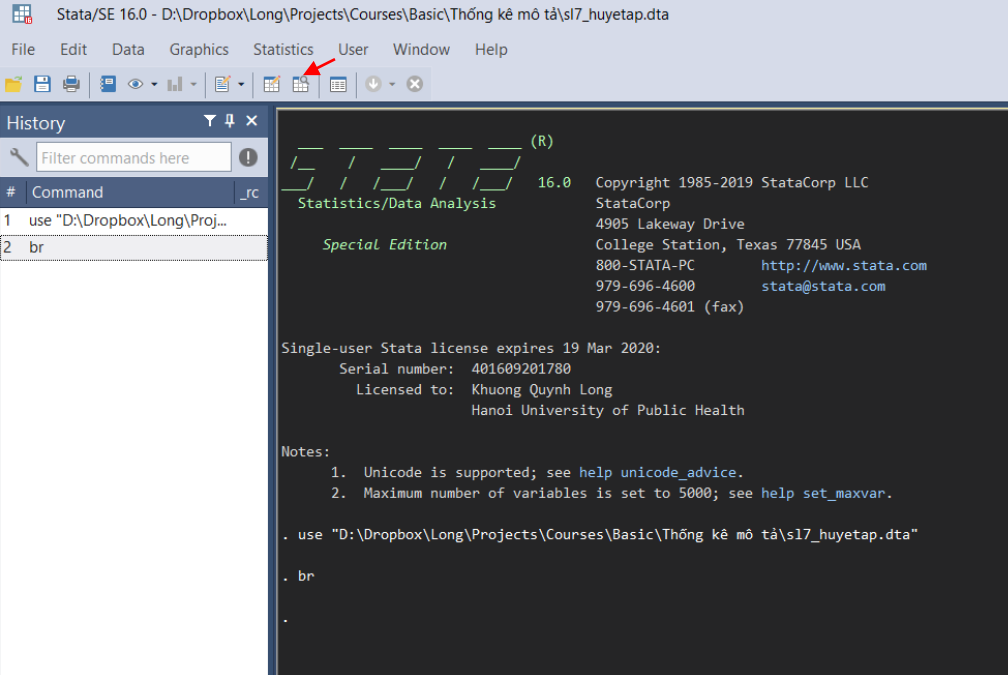
***Variable label:*** nhãn của biến số

***Sorted by:*** sắp xếp số liệu theo biến số nào

## 2. Xem cấu trúc số liệu

- Xem cấu trúc số liệu

br



The screenshot shows the Stata/SE 16.0 interface. The title bar indicates the file path: D:\Dropbox\Long\Projects\Courses\Basic\Thống kê mô tả\s17\_huyetap.dta. The menu bar includes File, Edit, Data, Graphics, Statistics, User, Window, and Help. The toolbar contains various icons for file operations and analysis. The History window on the left shows a list of commands: 1 use "D:\Dropbox\Long\Proj..." and 2 br. The main command window on the right displays the Stata startup screen, which includes the Stata logo, version 16.0, copyright information (1985-2019 StataCorp LLC), and the user's license details (Single-user Stata license expires 19 Mar 2020, Serial number: 401609201780, Licensed to: Khuong Quynh Long, Hanoi University of Public Health). The command history in the main window shows: . use "D:\Dropbox\Long\Projects\Courses\Basic\Thống kê mô tả\s17\_huyetap.dta" . br .

| Data Editor (Browse) - [si7_huyetap.dta] |    |      |          |         |        |           |             |              |           |           |          |          |             |             |          |          |
|--|----|------|----------|---------|--------|-----------|-------------|--------------|-----------|-----------|----------|----------|-------------|-------------|----------|----------|
| File Edit View Data Tools                |    |      |          |         |        |           |             |              |           |           |          |          |             |             |          |          |
| ma[1]                                    |    |      |          |         |        |           |             |              |           |           |          |          |             |             |          |          |
|  | ma | tuoi | gioitinh | cannang | caotb  | suckhoe   | ttthainghen | caoha        | chieucao1 | chieucao2 | hatoida1 | hatoida2 | hatoithieu1 | hatoithieu2 | nhomtuoi | hatdtb   |
| 1  | 2  | 49   | nu       | 38.4    | 154.6  | yeu       | khong       | khong cao ha | 154.6     | 154.6     | 90       | 90       | 60          | 45-54       | 60       | 16.06617 |
| 2  | 2  | 61   | nam      | 55.8    | 149.8  | tot       | khong       | cao ha       | 149.8     | 149.8     | 178      | 198      | 112         | 118         | 55-64    | 189      |
| 3  | 2  | 25   | nu       | 56.2    | 152.5  | binh thu. | khong       | khong cao ha | 152.5     | 152.5     | 105      | 105      | 70          | 70          | 25-34    | 106      |
| 4  | 2  | 60   | nu       | 57.2    | 162.5  | binh thu. | khong       | khong cao ha | 162.5     | 162.5     | 125      | 125      | 80          | 80          | 55-64    | 126      |
| 5  | 1  | 32   | nu       | 62.5    | 170.15 | binh thu. | .           | khong cao ha | 170.1     | 170.2     | 106      | 102      | 76          | 70          | 25-34    | 105      |
| 6  | 1  | 61   | nam      | 45.7    | 162.15 | binh thu. | .           | cao ha       | 162.2     | 162.1     | 170      | 166      | 100         | 104         | 55-64    | 169      |
| 7  | 1  | 34   | nam      | 46      | 162.1  | tot       | .           | khong cao ha | 162.1     | 162.1     | 115      | 115      | 70          | 70          | 25-34    | 116      |
| 8  | 1  | 45   | nu       | 63.1    | 167.35 | tot       | .           | khong cao ha | 167.3     | 167.4     | 115      | 120      | 65          | 60          | 45-54    | 118.5    |
| 9  | 1  | 40   | nu       | 52.4    | 166.5  | binh thu. | .           | khong cao ha | 166.5     | 166.5     | 100      | 100      | 65          | 65          | 35-44    | 101      |
| 10                                       | 3  | 26   | nu       | 58      | 170    | binh thu. | .           | khong cao ha | 170       | 170       | 120      | 120      | 80          | 80          | 25-34    | 121      |
| 11                                       | 2  | 48   | nu       | 40.9    | 151.7  | binh thu. | khong       | khong cao ha | 151.6     | 151.8     | 120      | 120      | 70          | 75          | 45-54    | 121      |
| 12                                       | 2  | 47   | nam      | 48.4    | 151.7  | yeu       | khong       | cao ha       | 151.6     | 151.8     | 130      | 122      | 98          | 95          | 45-54    | 127      |
| 13                                       | 3  | 31   | nu       | .       | .      | binh thu. | khong       | khong cao ha | .         | .         | 130      | 120      | 80          | 70          | 25-34    | 126      |
| 14                                       | 1  | 36   | nu       | 64.2    | 176    | tot       | .           | khong cao ha | 176       | .         | 120      | 120      | 80          | 80          | 35-44    | 121      |
| 15                                       | 3  | 30   | nu       | 44.4    | 148.95 | binh thu. | khong       | khong cao ha | 148.9     | 149       | 114      | 116      | 68          | 70          | 25-34    | 116      |
| 16                                       | 1  | 42   | nam      | 54.2    | 157.95 | binh thu. | .           | khong cao ha | 157.9     | 158       | 110      | 110      | 70          | 70          | 35-44    | 111      |
| 17                                       | 1  | 48   | nam      | 45      | 155.65 | tot       | .           | khong cao ha | 155.6     | 155.7     | 140      | 135      | 90          | 85          | 45-54    | 138.5    |
| 18                                       | 2  | 43   | nu       | 52.3    | 159.1  | binh thu. | khong       | khong cao ha | 159.1     | 159.1     | 120      | 120      | 70          | 70          | 35-44    | 121      |
| 19                                       | 5  | 44   | nu       | 38.9    | 150.1  | binh thu. | khong       | khong cao ha | 150.1     | 150.1     | 120      | 120      | 80          | 80          | 35-44    | 121      |
| 20                                       | 1  | 49   | nam      | 51.5    | 148.05 | binh thu. | khong       | cao ha       | 148.1     | 148       | 170      | 165      | 100         | 100         | 45-54    | 168.5    |
| 21                                       | 2  | 34   | nam      | 53.1    | 158.75 | binh thu. | khong       | khong cao ha | 158.7     | 158.8     | 120      | 115      | 80          | 75          | 25-34    | 118.5    |
| 22                                       | 1  | 44   | nam      | 52      | 160.1  | binh thu. | .           | khong cao ha | 160.1     | 160.1     | 110      | 110      | 60          | 60          | 35-44    | 111      |
| 23                                       | 2  | 51   | nam      | 52.6    | 159.95 | binh thu. | .           | khong cao ha | 159.9     | 160       | 120      | 130      | 70          | 70          | 45-54    | 126      |
| 24                                       | 2  | 28   | nu       | 41      | 154.3  | binh thu. | co thai     | khong cao ha | 155.3     | 153.3     | 110      | 110      | 75          | 75          | 25-34    | 111      |
| 25                                       | 4  | 26   | nam      | 46.9    | 157.5  | binh thu. | .           | khong cao ha | 157.5     | 157.5     | 120      | 120      | 90          | 88          | 25-34    | 121      |
| 26                                       | 2  | 29   | nu       | 48.9    | 160.9  | binh thu. | khong       | khong cao ha | 160.9     | 160.9     | 108      | 106      | 70          | 68          | 25-34    | 108      |
| 27                                       | 2  | 33   | nam      | 43.6    | 152.4  | yeu       | dang cho bu | khong cao ha | 152.5     | 152.3     | 118      | 116      | 80          | 78          | 25-34    | 118      |
| 28                                       | 1  | 44   | nam      | 62.1    | 165.65 | binh thu. | .           | khong cao ha | 165.6     | 165.7     | 126      | 120      | 84          | 80          | 35-44    | 124      |
| 29                                       | 2  | 46   | nu       | 53.9    | 150.5  | binh thu. | khong       | khong cao ha | 150.5     | 150.5     | 120      | 120      | 60          | 60          | 45-54    | 121      |
| 30                                       | 6  | 36   | nam      | 50.9    | 171.45 | binh thu. | .           | cao ha       | 171.5     | 171.4     | 140      | 138      | 100         | 100         | 35-44    | 140      |
| 31                                       | 1  | 42   | nu       | 53.5    | 165.35 | binh thu. | .           | khong cao ha | 165.3     | 165.4     | 120      | 115      | 80          | 75          | 35-44    | 118.5    |
| 32                                       | 8  | 37   | nu       | 45.2    | 154.5  | binh thu. | khong       | khong cao ha | 154.4     | 154.6     | 110      | 110      | 70          | 65          | 35-44    | 111      |
| 33                                       | 1  | 32   | nam      | 51.9    | 162.65 | rat tot   | .           | khong cao ha | 162.6     | 162.7     | 120      | 120      | 85          | 90          | 25-34    | 121      |
| 34                                       | 2  | 39   | nu       | 40.6    | 152.45 | binh thu. | khong       | khong cao ha | 152.5     | 152.4     | 120      | 120      | 80          | 80          | 35-44    | 121      |

Filter variables here

| Name  | Label                  |
|---|------------------------|
| <input checked="" type="checkbox"/> ma          | ma ca nhan             |
| <input checked="" type="checkbox"/> tuoi        | tuoi                   |
| <input checked="" type="checkbox"/> gioitinh    | gioi tinh              |
| <input checked="" type="checkbox"/> cannang     | can nang               |
| <input checked="" type="checkbox"/> caotb       | chieu cao tb           |
| <input checked="" type="checkbox"/> suckhoe     | nhan xet ve suc khoe   |
| <input checked="" type="checkbox"/> ttthainghen | trinh trang thai nghen |
| <input checked="" type="checkbox"/> caoha       | cao huyet ap           |
| <input checked="" type="checkbox"/> chieucao1   | chieu cao lan 1        |
| <input checked="" type="checkbox"/> chieucao2   | chieu cao lan 2        |
| <input checked="" type="checkbox"/> hatoida1    | huyet ap toi da 1      |

Variables Snapshots

Variables

|             |            |
|-------------|------------|
| Name        | ma         |
| Label       | ma ca nhan |
| Type        | float      |
| Format      | %9.0g      |
| Value label |            |
| Notes       |            |

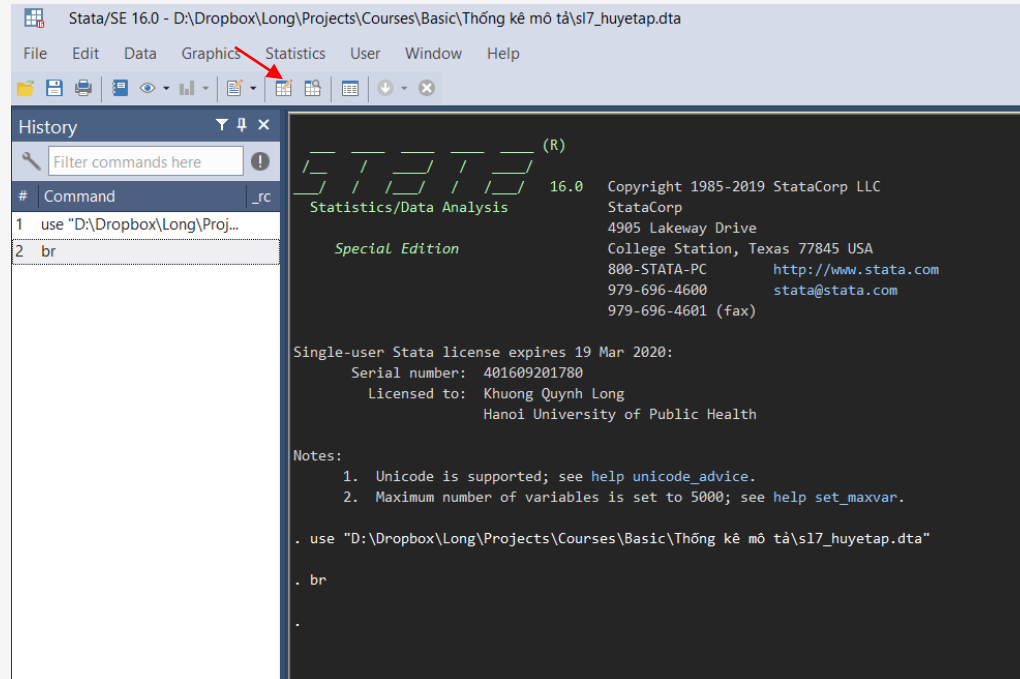
Data

|              |                 |
|--------------|-----------------|
| Frame        | default         |
| Filename     | si7_huyetap.dta |
| Label        |                 |
| Notes        |                 |
| Variables    | 19              |
| Observations | 397             |
| Size         | 29.46K          |
| Memory       | 64M             |
| Sorted by    |                 |

### 3. Tìm kiếm và chỉnh sửa số liệu

- Tìm kiếm và chỉnh sửa số liệu

edit



The screenshot shows the Stata/SE 16.0 interface. The title bar indicates the file path: D:\Dropbox\Long\Projects\Courses\Basic\Thống kê mô tả\s17\_huyetap.dta. The menu bar includes File, Edit, Data, Graphics, Statistics, User, Window, and Help. A red arrow points to the 'Statistics' menu. The left sidebar contains a 'History' panel with a search bar and a list of commands. The main command window displays the Stata startup screen, including the Stata logo, version 16.0, copyright information (1985-2019 StataCorp LLC), and the user's license details. The command history shows two commands: 'use "D:\Dropbox\Long\Proj...' and 'br'.

```
Stata/SE 16.0 - D:\Dropbox\Long\Projects\Courses\Basic\Thống kê mô tả\s17_huyetap.dta
File Edit Data Graphics Statistics User Window Help

History
Filter commands here
# Command _rc
1 use "D:\Dropbox\Long\Proj..."
2 br

(R)
-----
Statistics/Data Analysis
Special Edition

Copyright 1985-2019 StataCorp LLC
StataCorp
4905 Lakeway Drive
College Station, Texas 77845 USA
800-STATA-PC http://www.stata.com
979-696-4600 stata@stata.com
979-696-4601 (fax)

Single-user Stata license expires 19 Mar 2020:
Serial number: 401609201780
Licensed to: Khuong Quynh Long
Hanoi University of Public Health

Notes:
1. Unicode is supported; see help unicode_advice.
2. Maximum number of variables is set to 5000; see help set_maxvar.

. use "D:\Dropbox\Long\Projects\Courses\Basic\Thống kê mô tả\s17_huyetap.dta"

. br

.
```

## 4. Sắp xếp số liệu

---

- Sắp xếp theo giá trị tăng dần

`sort {tên biến cần sắp xếp}`

`sort tuoi`

`sort id`

- Sắp xếp theo giá trị giảm dần

`gsort - {tên biến cần sắp xếp}`

`gsort - tuoi`



## 5. Lưu số liệu

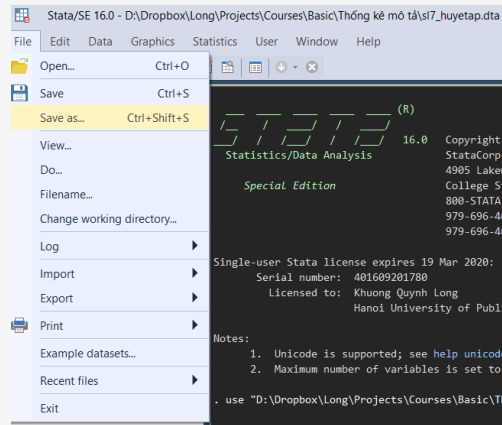
- Lưu file Stata

- ✓ Cách 1

**save {đường dẫn}\tendata.dta, replace**

- ✓ Cách 2

file >> save as >> chọn đường dẫn và đặt tên



## 6. Tạo biến mới

---

- Tạo biến mới

**gen {biến mới} = ...**

- Ví dụ:

```
gen chieucaotb = (chieucaol + chieucaol2)/2
```

```
gen bmi = kannang/chieucaol^2
```

```
gen tangha = hatdtb >= 140 | hatttb >= 90
```

Chú ý missing value!

## 7. Đổi tên, xóa biến số

---

- Đổi tên

**ren {tên cũ} {tên mới}**

ren caoha caoha\_moi

- Xóa biến số

**drop {tên biến}**

drop chieucaol chieucaol2

- Giữ lại biến số

**keep {tên biến}**

keep chieucaol chieucaol2

## 8. Tạo và gán nhãn

---

- Tạo nhãn giá trị

**label define {tên nhãn} # "nhãn" # "nhãn"...**

- ✓ Ví dụ: `label define lbcaoha 1 "Co" 0 "Khong"`

- Gán nhãn giá trị

**label value {tên biên} {tên nhãn}**

- ✓ Ví dụ: `label value caoha lbcaoha`

## 9. Thay thế giá trị số liệu

---

**replace {biến số} = ... if ...**

Ví dụ:

- ✓ `replace gioitinh = 0 if gioitinh == 2`
- ✓ `replace suckhoe = . if suckhoe == 999`

## 10. Phân nhóm biến số

---

**recode {biến số} (quy luật) , gen({biến mới})**

- Ví dụ: tạo biến nhóm BMI <18.5, 18.5 – 25, >=25

```
recode bmi (min/18.49999 = 1) (18.5/24.9999 = 2) (25/max = 3) , gen(bmi_nhom)
```

- Ví dụ: tạo biến nhóm BMI, kèm theo nhãn giá trị

```
recode bmi (min/18.49999 = 1 "<18.5") (18.5/24.9999 = 2  
"18.5 - <25") (25/max = 3 ">= 25") , gen(bmi_nhom)
```

# 11. Chuyển định dạng số liệu

|    | mhgd       | mcn          | age | edu |  |
|----|------------|--------------|-----|-----|--|
| 1  | 0101010001 | 010101000101 | 41  | 2   |  |
| 2  | 0101010002 | 010101000203 | 28  | 3   |  |
| 3  | 0101010003 | 010101000304 | 12  | 3   |  |
| 4  | 0101010004 | 010101000402 | 34  | 3   |  |
| 5  | 0101010005 | 010101000504 | 9   | 2   |  |
| 6  | 0101010006 | 010101000601 | 40  | 2   |  |
| 7  | 0101010008 | 010101000807 | 1   | 1   |  |
| 8  | 0101010009 | 010101000904 | 13  | 3   |  |
| 9  | 0101010010 | 010101001005 | 27  | 4   |  |
| 10 | 0101010013 | 010101001303 | 21  | 3   |  |

Dạng string

Dạng numeric

Chuyển từ string → numeric

**destring {tênbiến}, gen(biếnmới)**

Hoặc

**destring {tên biến}, replace**

**destring mcn, replace**

Chuyển từ numeric → string

**tostring {tênbiến}, gen(biếnmới)**

Hoặc

**tostring {tên biến}, replace**

**tostring age, replace**

## 12. Gộp bộ số liệu

- Gộp theo hàng dọc (append)
- Các dữ liệu phải có
  - ✓ Cùng tên biến
  - ✓ Cùng định dạng (string, numeric)
  - ✓ Nếu không cùng tên?

use "Cambodia.dta", clear  
append using "Angola.dta"

### Cambodia

|    | country | country_name | year | neonatal | infant | under5 |
|----|---------|--------------|------|----------|--------|--------|
| 1  | KH6     | Cambodia     | 2014 | No       | No     | No     |
| 2  | KH6     | Cambodia     | 2014 | No       | No     | No     |
| 3  | KH6     | Cambodia     | 2014 | No       | No     | No     |
| 4  | KH6     | Cambodia     | 2014 | No       | No     | No     |
| 5  | KH6     | Cambodia     | 2014 | No       | No     | No     |
| 6  | KH6     | Cambodia     | 2014 | No       | No     | No     |
| 7  | KH6     | Cambodia     | 2014 | No       | No     | No     |
| 8  | KH6     | Cambodia     | 2014 | No       | No     | No     |
| 9  | KH6     | Cambodia     | 2014 | No       | No     | No     |
| 10 | KH6     | Cambodia     | 2014 | No       | No     | No     |

### Angola

|    | country | country_name | year | neonatal | infant | under5 |
|----|---------|--------------|------|----------|--------|--------|
| 1  | A07     | Angola       | 2016 | No       | No     | No     |
| 2  | A07     | Angola       | 2016 | No       | No     | No     |
| 3  | A07     | Angola       | 2016 | No       | No     | No     |
| 4  | A07     | Angola       | 2016 | No       | No     | No     |
| 5  | A07     | Angola       | 2016 | No       | No     | No     |
| 6  | A07     | Angola       | 2016 | No       | No     | No     |
| 7  | A07     | Angola       | 2016 | No       | No     | No     |
| 8  | A07     | Angola       | 2016 | No       | No     | No     |
| 9  | A07     | Angola       | 2016 | No       | No     | No     |
| 10 | A07     | Angola       | 2016 | No       | No     | No     |



## 12. Gộp bộ số liệu

- Gộp theo hàng ngang

Phiếu hộ gia đình + cá nhân

|    | mhgd       | mcn          | eth | sex    | edu        |
|----|------------|--------------|-----|--------|------------|
| 1  | 0101010001 | 010101000101 | Dao | Male   | Primary    |
| 2  | 0101010002 | 010101000203 | Dao | Male   | Secondary  |
| 3  | 0101010003 | 010101000304 | Dao | Female | Secondary  |
| 4  | 0101010004 | 010101000402 | Dao | Female | Secondary  |
| 5  | 0101010005 | 010101000504 | Dao | Male   | Primary    |
| 6  | 0101010006 | 010101000601 | Dao | Male   | Primary    |
| 7  | 0101010008 | 010101000807 | Dao | Male   | Illiterate |
| 8  | 0101010009 | 010101000904 | Dao | Female | Secondary  |
| 9  | 0101010010 | 010101001005 | Dao | Female | High+      |
| 10 | 0101010013 | 010101001303 | Dao | Female | Secondary  |

Phiếu số 2

|    | mcn          | anc4 | inst_delivery | tetanus |
|----|--------------|------|---------------|---------|
| 1  | 010101000204 | No   | Yes           | Yes     |
| 2  | 010101000402 | No   | Yes           | Yes     |
| 3  | 010101000805 | Yes  | Yes           | Yes     |
| 4  | 010101001005 | No   | Yes           | Yes     |
| 5  | 010102000102 | No   | Yes           | Yes     |
| 6  | 010102000401 | No   | No            | No      |
| 7  | 010102000602 | Yes  | Yes           | Yes     |
| 8  | 010102001002 | No   | Yes           | Yes     |
| 9  | 010103000202 | Yes  | Yes           | Yes     |
| 10 | 010103000402 | No   | Yes           | Yes     |

## 12. Gộp bộ số liệu

---

- Có 3 dạng chính
  - ✓ **1:1 Ghép 1 dòng với 1 dòng**
    - Ví dụ: mcn ở phiếu 2 và mcn ở phiếu hộ gia đình
  - ✓ **1:m (hoặc m:1) Ghép 1 dòng với nhiều dòng**
    - Ví dụ: mghd ở phiếu hộ gia đình vs mhgd ở phiếu cá nhân
  - ✓ **m:m Ghép nhiều dòng với nhiều dòng**

## 12. Gộp bộ số liệu

```
use "phieuhgd.dta", clear
```

```
merge 1:1 mcn using "phieu2.dta"
```

| Result      | # of obs.         |
|-------------|-------------------|
| not matched | 6,203             |
| from master | 0 (_merge==1)     |
| from using  | 6,203 (_merge==2) |
| matched     | 966 (_merge==3)   |

Số mẫu không matched

- Có ở data đang mở mà không có ở data merge
- Có ở data merge mà không có ở data đang mở

Số mẫu matched (có ở cả 2 data)

## 13. preserve và restore

---

- Tạo ra 1 môi trường làm việc tạm thời
- Khi cần phân tích trên bộ số liệu tạm thời mà không làm ảnh hưởng tới dữ liệu đang sử dụng

**preserve** (bắt đầu tạo môi trường làm việc tạm thời)

Thực thi các câu lệnh

**restore** (trở về môi trường mặc định)

## 12. preserve và restore

\* Hôn nhân cận huyết thống

preserve

keep if chuho==1 & age\_cat>2

table1, vars(hncht cat\) ///

format(%9.1f) onecol ///

saving("4\_hncht\_c.xls", sheet (1, replace))

basetable hncht eth (r) db (r) xlkt (r) , toxl(4\_hncht\_r.xls, hncht, replace)

restore

\* Yếu tố hành vi ở người trưởng thành

preserve

keep if age\_cat>2

table1, vars(smoke cat\ drink cat\ smoke\_drink cat\) ///

format(%9.1f) onecol ///

saving("4\_ythv\_c.xls", sheet (1, replace))

basetable smoke eth (r) db (r) xlkt (r) , toxl(4\_ythv\_r.xls, hutthuoc, replace)

basetable drink eth (r) db (r) xlkt (r) , toxl(4\_ythv\_r.xls, uongruou, replace)

basetable smoke\_drink eth (r) db (r) xlkt (r) , toxl(4\_ythv\_r.xls, thuoc&ruou, replace)

restore

## 13. Loops

---

- Giả sử cần lặp lại thao tác giống nhau cho 10, 100, 1000... biến. Thao tác bằng cách copy & paste?
  - ✓ Dễ xảy ra lỗi
  - ✓ Khi sửa, cần sửa hàng loạt
  - ✓ Câu lệnh dài, khó đọc
- ➔ Loop
- **foreach** loop
  - **forvalue** loop

# Loops

---

- foreach loop

```
foreach macroname in list {  
    câu lệnh liên quan đến macroname  
}
```

```
foreach macroname of list-type {  
    câu lệnh liên quan đến macroname  
}
```

- List-type bao gồm:
  - ✓ **local**, **global**
  - ✓ **varlist**, numlist, newlist

# Loops

```
foreach var of varlist age weight height sex overweight {  
    logistic hypertension `var'  
}
```

Đặt tên bất kỳ





# Loop

---

```
240
241 ***** Convert text into number *****
242 foreach var of varlist Q17-Q101 {
243     replace `var' = "" if `var' == "."
244     replace `var' = "" if `var' == "*"
245     replace `var' = "1" if `var' == "A"
246     replace `var' = "2" if `var' == "B"
247     replace `var' = "3" if `var' == "C"
248     replace `var' = "4" if `var' == "D"
249     replace `var' = "5" if `var' == "E"
250     replace `var' = "6" if `var' == "F"
251     replace `var' = "7" if `var' == "G"
252     replace `var' = "8" if `var' == "H"
253     replace `var' = "9" if `var' == "I"
254 }
255
256 * destring
257 destring province-Q101, replace
```

# Loop

```
621 ***** Bảng 1: Thông tin chung
622 table1, vars(eth cat\ sex cat\ age_cat cat\ edu cat\ nn cat\ job cat\ religion cat\ xlkt cat\) ///
623 format(%9.1f) onecol ///
624 saving("ncd1.xls", sheet (1, replace))
625
626 ***** Bảng 2: Tỷ lệ mắc bệnh ncd
627 table1, vars(cvd cat\ dia cat\ cancer cat\ copd cat\ mental cat\ ncd_yn cat\) ///
628 format(%9.1f) onecol ///
629 saving("ncd2.xls", sheet (1, replace))
630
631 foreach var of varlist cvd dia cancer copd ncd_yn {
632     ci proportions `var'
633 }
634
635 foreach var of varlist eth sex age_cat edu nn job religion xlkt {
636     mean ncd_yn, over(`var')
637 }
638
```