SEMINAR

TẠO BẢNG BẰNG STATA

Đại Học Y Tế Công Cộng

Khương Quỳnh Long
Hà Nội, 23/11/2019

Nội dung

1. Giới thiệu Stata macros
   - Loops
2. Tạo các bảng cho bài báo
3. Quản lý thư mục

Stata macros & loops

Stata macros

- Macros là các ký tự (tên) viết tắt, đại diện cho một chuỗi các ký tự, số, tên biến hoặc biểu thức, giúp lưu trữ hoặc rút ngắn thời gian sao chép

```
desc age weight height bmi systolic_BP diastolic_BP
mdesc age weight height bmi systolic_BP diastolic_BP
sum age weight height bmi systolic_BP diastolic_BP
pwcorr age weight height bmi systolic_BP diastolic_BP
```

X

```
global vars age weight height bmi systolic_BP diastolic_BP
desc $vars
mdesc $vars
sum $vars
pwcorr $vars
```

Stata macros

- "**global**" & "**local**"
- **local** : tạm thời, chỉ có tác dụng trong lần thao tác đó
- **global**: tồn tại đến khi nào xóa bỏ hoặc kết thúc session
- Sử dụng:

✓ **local** macroname "string"/variable

✓ **local** macroname = expresion

✓ **global** macroname "string"/variable

✓ **global** macroname = expresion

Stata macros

- local: `macroname'
- global: $macroname
- Ví dụ:

```
global myvars age weight height bmi systolic_BP
sum $myvars
```

- myvars "đại diện" cho 5 biến phía sau, khi cần sử dụng 5 biến này, chỉ cần dùng $myvars

```
global mydir D:\Dropbox\Long\Projects\Courses\Writing
save "$myvars/data.dta", replace
```

- mydir "đại diện" cho đường dẫn, sử dụng $mydir để thay cho đường dẫn dài ➔ hữu ích để quản lý thư mục

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:

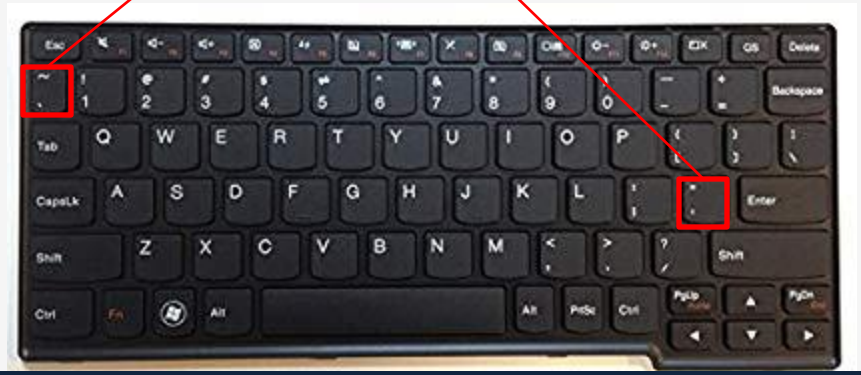
✓ **local**, **global**

✓ **varlist**, numlist, newlist

Loops

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

Đặt tên bất kỳ



```
logistic hypertension age
logistic hypertension weight
logistic hypertension height
logistic hypertension sex
logistic hypertension overweight
```

Loops

- Kết hợp với local hoặc global

```
global myvar age weight height sex overweight
sum $myvar

foreach var of varlist $myvar {
        logistic hypertension `var'
}
```

Tạo các bảng với Stata

1

Variables	Category	Frequency	%
Age	≤ 30 years	124	27.2
	31–40 years	111	24.3
	41–50 years	91	20.0
	> 50 years	130	71.5
Sex	Male	208	45.6
	Female	248	54.4
Marital status	Single	47	10.3
	Married	348	76.3
	Widowed	29	6.4
	Divorced	32	7.0
Educational status of respondent	Cannot read and write	92	20.2
	Can read and write	110	24.1
	Grade 1–8	150	32.9
	Grade 9–12	38	8.3
	Certificates and above	66	14.5
Religion of respondents	Orthodox	442	96.9
	Muslim	14	3.1
Ethnicity of respondents	Amhara	449	98.5
	Oromo	6	1.3
	Tigray	1	0.2
Occupational status	Employed	127	27.9
	Merchant	184	40.4
	Housewife	52	11.4
	Student	11	2.4
	Others	82	18.0

2

Demographic	Unsafe sexual practices		
	Yes *n* (%)	No *n* (%)	*p*
Sex			
Female	25 (17.1)	121 (82.9)	0.137
Male	30 (11.8)	224 (88.2)	
Sexual orientation			
Heterosexual	51 (14.5)	300 (85.5)	0.370
Homosexual/Bisexual	3 (13.0)	20 (87.0)	
Unsure/Not answered	1 (3.8)	25 (96.2)	
Age in years			
≤ 30	21 (21.0)	79 (79.0)	0.035
31–40	29 (12.2)	208 (87.8)	
≥ 40	5 (7.9)	58 (92.1)	
Median (IQR)	31 (30–37)	34 (31–38)	0.022
Work status			
Unemployed/Casual	21 (13.8)	131 (86.2)	0.893
Part-time	9 (13.0)	60 (87.0)	
Full-time	16 (12.7)	110 (87.3)	
Not working	9 (17.0)	44 (83.0)	
Highest level of education completed			
≤ Primary school	20 (14.4)	119 (85.6)	0.030
Secondary school	13 (8.6)	138 (91.4)	
≥ High school	22 (20.0)	88 (80.0)	

3

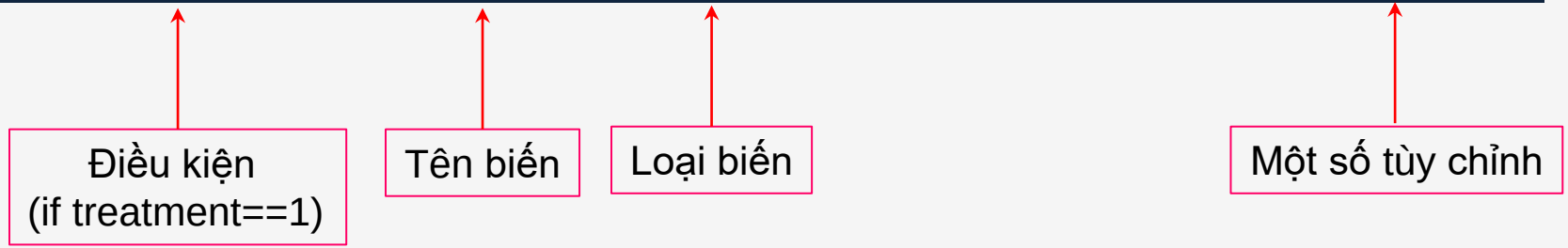
Predictor	Odds Ratio	95% CI	*P* Value
Sex (female)	1.77	1.06-2.96	.029
Age (years)	1.05	1.02-1.09	.003
Education			
≤ Primary school	References	–	–
Secondary school	0.83	0.49-1.40	.484
≥ High school	0.55	0.30-0.99	.046
Source of HIV infection			
Sexual transmission	References	–	–
Injected drug use	1.53	0.89-2.64	.126
Others	3.57	1.82-7.01	<.001
HIV stage			
1	References	–	–
2	3.14	0.81-12.08	.097
3	3.60	1.16-11.19	.027
4	2.76	0.38-19.87	.314
Adherence to HIV medication			
Good	References	–	–
Moderate	3.57	1.17-10.89	.025
Poor	1.19	0.36-3.93	.772
Depressive symptoms (yes)	3.23	2.05-5.11	<.001

“Table 1”

```
ssc install table1
```

- Sử dụng lệnh “`table1`” trong Stata

```
table1 [if], vars(varname1 vartype1\varname2 vartype3…) [options]
```



Vartype:

- continuous, normally distributed [Mean (SD)] ➔ **`contn`**
- continuous, skew [Median (IQR)] ➔ **`conts`**
- Categorical [n (%)] ➔ **`cat/cate`**

Option

- Điều chỉnh số thập phân
- Save
- Hiển thị cột

```
table1, vars(age contn\ age conts\ sex cat\ occupation cat\
             marital cat\ weight contn\ height contn)
```

Factor	Level	Value
N		397
Age, year, mean (SD)		41.68262 (10.58372)
Age, year, median (IQR)		41 (33, 49)
Sex	Male	201 (50.6%)
	Female	196 (49.4%)
Occupation	Unemployed	90 (22.7%)
	Worker	206 (51.9%)
	Officer	101 (25.4%)
Marital status	Single	108 (27.2%)
	Married	196 (49.4%)
	Windowed/separated	93 (23.4%)
Weight, kg, mean (SD)		51.37683 (8.240012)
Height, cm, mean (SD)		157.3957 (7.690895)

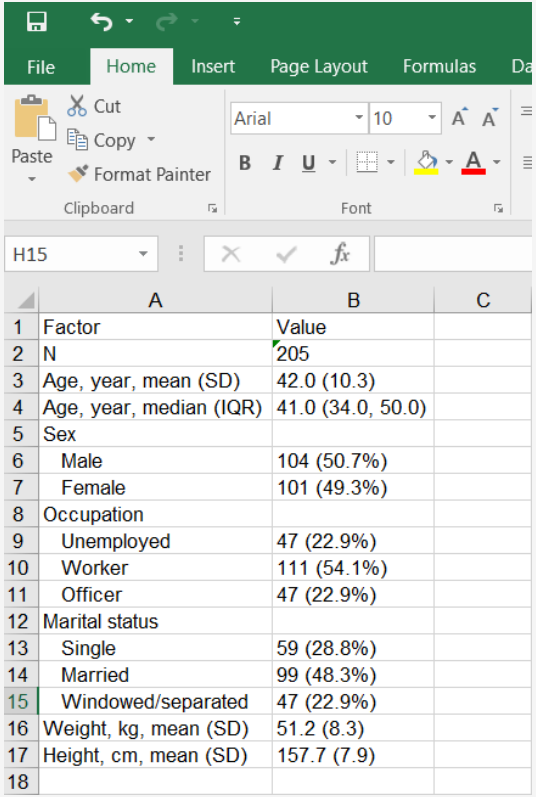
Format

```
table1, vars(age contn\ age conts\ sex cat\ occupation cat\ ///
             marital cat\ weight contn\ height contn) ///
             format(%9.1f) cformat(%9.1f) onecol
```

Factor	Value
N	397
Age, year, mean (SD)	41.7 (10.6)
Age, year, median (IQR)	41.0 (33.0, 49.0)
Sex	
Male	201 (50.6%)
Female	196 (49.4%)
Occupation	
Unemployed	90 (22.7%)
Worker	206 (51.9%)
Officer	101 (25.4%)
Marital status	
Single	108 (27.2%)
Married	196 (49.4%)
Windowed/separated	93 (23.4%)
Weight, kg, mean (SD)	51.4 (8.2)
Height, cm, mean (SD)	157.4 (7.7)

Số thập phân (1)

```
table1 if treatment == 1, vars(age contn\ age conts\ sex cat\ occupation cat\ ///
                               marital cat\ weight contn\ height contn) ///
                               format(%9.1f) cformat(%9.1f) onecol ///
                               saving("D:\Paper\Desc.xls", sheet (Table1, replace))
```



	A	B	C
1	Factor	Value	
2	N	205	
3	Age, year, mean (SD)	42.0 (10.3)	
4	Age, year, median (IQR)	41.0 (34.0, 50.0)	
5	Sex		
6	Male	104 (50.7%)	
7	Female	101 (49.3%)	
8	Occupation		
9	Unemployed	47 (22.9%)	
10	Worker	111 (54.1%)	
11	Officer	47 (22.9%)	
12	Marital status		
13	Single	59 (28.8%)	
14	Married	99 (48.3%)	
15	Windowed/separated	47 (22.9%)	
16	Weight, kg, mean (SD)	51.2 (8.3)	
17	Height, cm, mean (SD)	157.7 (7.9)	
18			

Điều kiện

Đường dẫn và tên file

Tên excel sheet

Cho phép ghi đè nếu chạy lại code

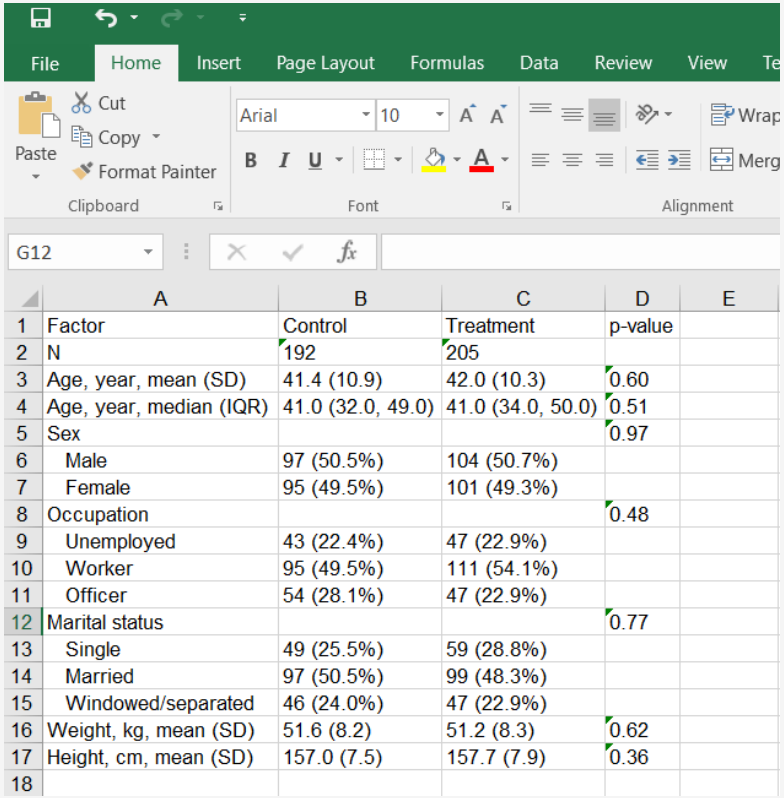
“Table 2”

```
table1, by(treatment) vars(age contn\ age conts\ sex cat\ occupation cat\ ///
                           marital cat\ weight contn\ height contn) ///
                           format(%9.1f) cformat(%9.1f) onecol ///
                           saving("D:\Paper\Desc.xls", sheet (Table2, replace))
```

by (“group”)

Lựa chọn kiểm định: tùy thuộc “vartype”

- **contn** ➔ ANOVA
- **conts** ➔ Wilcoxon rank-sum (2 groups) or Kruskal-Wallis (>2 groups)
- **cat** ➔ Chi-squared
- **cate** ➔ Fisher's exact



	A	B	C	D	E
1	Factor	Control	Treatment	p-value	
2	N	192	205		
3	Age, year, mean (SD)	41.4 (10.9)	42.0 (10.3)	0.60	
4	Age, year, median (IQR)	41.0 (32.0, 49.0)	41.0 (34.0, 50.0)	0.51	
5	Sex			0.97	
6	Male	97 (50.5%)	104 (50.7%)		
7	Female	95 (49.5%)	101 (49.3%)		
8	Occupation			0.48	
9	Unemployed	43 (22.4%)	47 (22.9%)		
10	Worker	95 (49.5%)	111 (54.1%)		
11	Officer	54 (28.1%)	47 (22.9%)		
12	Marital status			0.77	
13	Single	49 (25.5%)	59 (28.8%)		
14	Married	97 (50.5%)	99 (48.3%)		
15	Windowed/separated	46 (24.0%)	47 (22.9%)		
16	Weight, kg, mean (SD)	51.6 (8.2)	51.2 (8.3)	0.62	
17	Height, cm, mean (SD)	157.0 (7.5)	157.7 (7.9)	0.36	
18					

“Table 3”

- Hồi quy đa biến

```
logistic hypertension age i.sex i.occupation i.marital bmi i.health_status i.treatment
```

hypertension	Odds Ratio	Std. Err.	z	P>|z|	[95% Conf.	Interval]
age	1.062685	.0149645	4.32	0.000	1.033756	1.092423
sex						
Female	1.380531	.3834514	1.16	0.246	.8009762	2.37943
occupation						
Worker	2.296147	.8852638	2.16	0.031	1.078511	4.888489
Officer	1.904412	.8159259	1.50	0.133	.8223792	4.410111
marital						
Married	1.193165	.4012273	0.53	0.599	.6172564	2.306403
Windowed/separated	1.217568	.4761598	0.50	0.615	.5657311	2.620453
bmi	1.186642	.059286	3.43	0.001	1.075952	1.30872
health_status						
Normal	1.073799	.3958797	0.19	0.847	.521325	2.211756
Bad	2.844281	1.330267	2.24	0.025	1.137278	7.113419
treatment						
Treatment	.9769935	.2689852	-0.08	0.933	.5695599	1.675884
_cons	.0001466	.0002072	-6.24	0.000	9.17e-06	.0023421

Note: _cons estimates baseline odds.

- Lưu mô hình
- Trích xuất thông tin ra MS Word

```
eststo: logistic hypertension age i.sex i.occupation i.marital bmi i.health_status i.treatment

        esttab using "D:\Paper\Table3.rtf", b(2) ci(2) replace ///
        wide label nogaps star( * 0.05 ** 0.01 *** 0.001) stats(r2 N) eform
        eststo clear
```

	(1) Hypertension	
Hypertension		
Age, year	1.06***	[1.03,1.09]
Male	1.00	[1.00,1.00]
Female	1.38	[0.80,2.38]
Unemployed	1.00	[1.00,1.00]
Worker	2.30*	[1.08,4.89]
Officer	1.90	[0.82,4.41]
Single	1.00	[1.00,1.00]
Married	1.19	[0.62,2.31]
Windowed/separated	1.22	[0.57,2.62]
Body Mass Index, kg/m2	1.19***	[1.08,1.31]
Good	1.00	[1.00,1.00]
Normal	1.07	[0.52,2.21]
Bad	2.84*	[1.14,7.11]
Control	1.00	[1.00,1.00]
Treatment	0.98	[0.57,1.68]
r2		
N	397.00	

Exponentiated coefficients; 95% confidence intervals in brackets

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Hiển thị hệ số exp (OR, RR…)

Đường dẫn và tên file

- Đơn biến

```
eststo: logistic hypertension age
eststo: logistic hypertension i.age_gr
eststo: logistic hypertension i.sex
eststo: logistic hypertension i.occupation
eststo: logistic hypertension i.marital
eststo: logistic hypertension weight
eststo: logistic hypertension height
eststo: logistic hypertension bmi
eststo: logistic hypertension overweight
eststo: logistic hypertension i.health_status
eststo: logistic hypertension i.treatment

esttab using "D:\Paper\Table3_db.rtf", b(2) ci(2) replace ///
wide label nogaps star( * 0.05 ** 0.01 *** 0.001) stats(r2 N) eform
eststo clear
```

	(1) Hypertension		(2) Hypertension		(3) Hypertension		(4) Hypertension		(5) Hypertension		(6) Hypertension		(7) Hypertension		(8) Hypertension		(9) Hypertension		(10) Hypertension		(11) Hypertension	
Hypertension																						
Age, year	1.07***	[1.05,1.10]																				
25-34			1.00	[1.00,1.00]																		
35-44			2.77*	[1.17,6.57]																		
45-54			5.30***	[2.29,12.26]																		
55-64			9.62***	[3.92,23.66]																		
Male					1.00	[1.00,1.00]																
Female					1.52	[0.92,2.50]																
Unemployed							1.00	[1.00,1.00]														
Worker							1.82	[0.91,3.63]														
Officer							1.71	[0.79,3.70]														
Single									1.00	[1.00,1.00]												
Married									1.20	[0.66,2.20]												
Windowed/separated									1.20	[0.59,2.44]												
Weight, kg											1.04*	[1.01,1.07]										
Height, cm													1.00	[0.96,1.03]								
Body Mass Index, kg/m2															1.15**	[1.05,1.26]						
Overweight																	3.41***	[1.95,5.97]				
Good																			1.00	[1.00,1.00]		
Normal																			0.94	[0.49,1.83]		
Bad																			3.05**	[1.39,6.69]		
Control																					1.00	[1.00,1.00]
Treatment																					0.98	[0.60,1.61]
r2																						
N	397.00		397.00		397.00		397.00		397.00		397.00		397.00		397.00		397.00		397.00		397.00	

Exponentiated coefficients; 95% confidence intervals in brackets
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

1

```
* biến phân loại
foreach var of varlist age_gr sex occupation marital health_status {
          eststo: logistic hypertension i.`var'
}
* biến liên tục
foreach var of varlist age weight height bmi overweight {
          eststo: logistic hypertension `var'
}
```

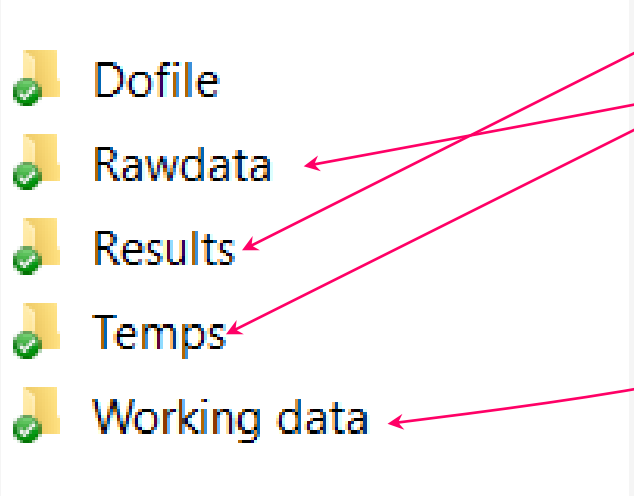
2

```
* Chọn tất cả các biến cần kiểm tra (dùng global)
global indvar age age_gr sex occupation marital weight height bmi overweight ///
             health_status treatment

foreach var of varlist $indvar {
          distinct `var'
          local x = r(ndistinct)

          if `x' < 10 & `x' != 2 {
                    eststo: logistic hypertension i.`var'
          }
          if `x' > 10 | `x' == 2 {
                    eststo: logistic hypertension `var'
          }
}
```

Quản lý thư mục

Thay thế đường dẫn dài bằng cách sử dụng global

```
global rawdata D:\Dropbox\Long\Projects\.....\Stata\Rawdata
global workdata D:\Dropbox\Long\Projects\.....\Stata\Working data
global result D:\Dropbox\Long\Projects\.....\Stata\Results
global temp D:\Dropbox\Long\Projects\.....\Stata\Temps
```



Thay thế đường dẫn dài bằng cách sử dụng global

```
use "$rawdata\Hypertension.dta", clear

table1 if treatment == 1, vars(age contn\ age conts\ sex cat\ occupation cat\ ///
                            marital cat\ weight contn\ height contn) ///
                            format(%9.1f) cformat(%9.1f) onecol ///
                            saving("$temp\Desc.xls", sheet (Table1, replace))

eststo: logistic hypertension age i.sex i.occupation i.marital bmi i.health_status
esttab using "$result\Table3.rtf", b(2) ci(2) replace ///
             wide label nogaps star( * 0.05 ** 0.01 *** 0.001) stats(r2 N) eform
eststo clear
```

Thank you!