Chương Trình Giảng Dạy Kinh tế Fulbright

Học kỳ Thu năm 2010

Các Phương Pháp Phân Tích Định Lượng

Gợi ý trả lời - Bài tập 10
LỰA CHỌN DẠNG HÀM & KIỂM ĐỊNH ĐẶC TRƯNG MÔ HÌNH
Ngày Phát: Thứ Hai, 20/12/2010
Ngày Nộp: **8:20 sáng**, **Thứ Hai, 03/01/2011**
Bản in nộp tại Phòng Giáo Vụ
Bản điện tử gởi đến thầy Nguyễn Khánh Duy theo địa chỉ
duynk@fetp.vnn.vn

Bài 1

Dữ liệu ở file **data 6-4.wf1** của Ramanathan, có các biến sau:

Wage: lương tháng của người lao động (USD)

Educ: Số năm đi học sau lớp tám (năm)

Exper: kinh nghiệm, đo lường bởi số năm đi học (năm)

Age: tuổi của người lao động (tuổi)

Bạn hãy thực hiện các thao tác cần thiết trên Eviews để tính toán những kết quả cần thiết nhằm trả lời các câu hỏi sau:

a. Ước lượng hàm hồi quy

$$Ln(wage)=\beta_1+\beta_2 educ+\boldsymbol{\beta_3} exper+\beta_4 age+ \boldsymbol{\beta_5} educ^2+\beta_6 exper^2+\beta_7 age^2 +u \qquad (1)$$

Sử dụng Eviews, bạn ước lượng được hàm hồi quy như sau (Xem Hình 1.1):

$$\widehat{Ln(wage)}=7.329-0.093educ+0.014exper-0.000426age+0.011525educ^2+0.000429exper^2+0.000021age^2$$

b. Giả sử rằng, theo lý thuyết, mô hình tổng quát về các yếu tố ảnh hưởng đến ln(wage) của người lao động như mô hình (1). Bạn hãy áp dụng phương pháp Hendry/LSE để tìm ra mô hình mà bạn cho là phù hợp nhất. Hãy giải thích vì sao bạn lại chọn mô hình đó.

Mô hình 1 (mô hình U) còn nhiều hệ số hồi quy không có ý nghĩa thống kê (ở độ tin cậy 90%), bạn lần lượt loại những biến không có ý nghĩa thống kê ra khỏi mô hình và mô hình mà các biến đều có ý nghĩa thống kê có thể là mô hình 2 (mô hình R)

$$\widehat{Ln(wage)}= 7.023+0.024exper+0.005educ^2$$

(Bạn cũng có thể chọn mô hình R là mô hình có biến exper, educ, $educ^2$ vì trong mô hình này có biến exper có ý nghĩa thống kê, và ít nhất một trong 2 biến educ, $educ^2$ có ý nghĩa thống kê; đặc biệt là biến $educ^2$ có ý nghĩa thống kê…)

Hình 1.1

Dependent Variable: LOG(WAGE)
Method: Least Squares
Date: 01/07/11 Time: 02:04
Sample: 1 49
Included observations: 49

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	7.329325	0.80918	9.058	0.000
EDUC	-0.093041	0.08639	-1.077	0.288
EXPER	**0.013863**	0.02448	0.566	0.574
AGE	-0.000426	0.03382	-0.013	0.990
EDUC^2	**0.011525**	0.00627	1.837	0.073
EXPER^2	0.000429	0.00112	0.384	0.703
AGE^2	0.000021	0.00038	0.055	0.956

R-squared	0.380615	Mean dependent var	7.454952
Adjusted R-squared	0.292131	S.D. dependent var	0.312741
S.E. of regression	0.263124	Akaike info criterion	0.299183
Sum squared resid	2.907844	Schwarz criterion	0.569443
Log likelihood	-0.329981	Hannan-Quinn criter.	0.401719
F-statistic	4.301529	Durbin-Watson stat	1.836202
Prob(F-statistic)	0.001816		

Hình 1.2

Dependent Variable: LOG(WAGE)
Method: Least Squares
Date: 01/07/11 Time: 02:17
Sample: 1 49
Included observations: 49

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	7.023367	0.092457	75.96326	0.0000
EXPER	0.023681	0.00614	3.856595	**0.0004**
EDUC^2	0.005023	0.001171	4.289093	0.0001

R-squared	0.361174	Mean dependent var	7.454952
Adjusted R-squared	0.333398	S.D. dependent var	0.312741
S.E. of regression	0.255339	Akaike info criterion	0.166823
Sum squared resid	2.999115	Schwarz criterion	0.282649
Log likelihood	-1.087164	Hannan-Quinn criter.	0.210767
F-statistic	13.00352	Durbin-Watson stat	1.691273
Prob(F-statistic)	0.000033		

Chúng ta nên thực hiện kiểm định Wald để kiểm định giả thuyết cho rằng các hệ số β_2 , β_4 , β_6 , β_7 đồng thời bằng không là đúng hay sai.

H_o: $\beta_2 = \beta_4 = \beta_6 = \beta_7 = 0$

H_1: Ít nhất một trong bốn hệ số trên khác không

P-value(F)=0.85 (>0.05) nên chưa đủ cơ sở để bác bỏ Ho ở độ tin cậy 95%. Nói cách khác, có thể chấp nhận Ho. Vì vậy, bạn có thể chọn mô hình sau là mô hình phù hợp

$\widehat{Ln(wage)} = 7.023+0.024exper+0.005educ^2$

Hình 1.3

Wald Test:
Equation: EQ01

Test Statistic	Value	df	Probability
F-statistic	0.3296	(4, 42)	**0.8565**
Chi-square	1.3183	4.0000	0.8583

Null Hypothesis Summary:

Normalized Restriction (= 0)	Value	Std. Err.
C(2)	-0.0930	0.0864
C(4)	-0.0004	0.0338
C(6)	0.0004	0.0011
C(7)	0.0000	0.0004

Restrictions are linear in coefficients.

c. Bạn hãy ước lượng hàm hồi quy sau:

$Ln(Wage)=\alpha_1+ \alpha_2 educ^2+ \alpha_3 exper + u$ (2)

Và cho biết, ý nghĩa kinh tế của hệ số $\hat{\alpha}_3$

Mô hình này tương tự như câu b. Hệ số 0.024 cho ta biết: trong điều kiện các yếu tố khác không đổi, khi exper tăng thêm 1 năm thì trung bình wage tăng lên 2.37 %.

d. Với mô hình (2), tại mức trung bình của educ và exper, bạn hãy tính tác động biên, và hệ số co giãn của wage theo exper

Ta tính được trung bình của exper=8.8367 và trung bình của educ=6.2245. Tại điểm này, wage=1681.165

Xét tại tại mức trung bình của educ và exper, với mô hình 2:

-Tác động cận biên của wage theo exper là $\hat{\alpha}_3$ wage=0.0237x1681.165=39.811

-Hệ số co giãn của wage theo exper là $\hat{\alpha}_3$ exper=0.0237x8.837=0.209

e. Bạn hãy cho biết với dạng hàm như mô hình (2), tại mức trung bình của educ và exper, tác động biên của educ lên wage là bao nhiêu? Và hệ số co giãn của wage theo educ là bao nhiêu?

Xét tại tại mức trung bình của educ và exper, với mô hình 2:

-Tác động cận biên của wage theo educ là

Wage.2. $\hat{\alpha}_2$ .educ= 1681.165x2x0.005x6.225=105.125

-Hệ số co giãn của wage theo educ là

Educ.2. $\hat{\alpha}_2$ .educ=6.225x2x0.005x6.225=0.389

Bài 2

Bạn hãy thực hiện các thao tác cần thiết trên Eviews để tính toán những kết quả cần thiết nhằm trả lời các câu hỏi a, b, c, d.

Mô hình 2 có dạng: $Ln(Wage)=\alpha_1+ \alpha_2 educ^2+ \alpha_3 exper +u$

a. Bạn hãy thực hiện kiểm định sai số đặc trưng của mô hình 2 bằng kiểm định Reset của Ramsey.

Từ cửa sổ Equation của mô hình 2, bạn chọn **View\Stability Test\Ramsey RESET Test,** sau đó nhập vào số 2 (khi đó mô hình hồi quy phụ sẽ có thêm 2 biến $\hat{Y}_i^2$ và $\hat{Y}_i^3$)

Kết quả ở Hình 2.2 cho phép ta thực hiện kiểm định

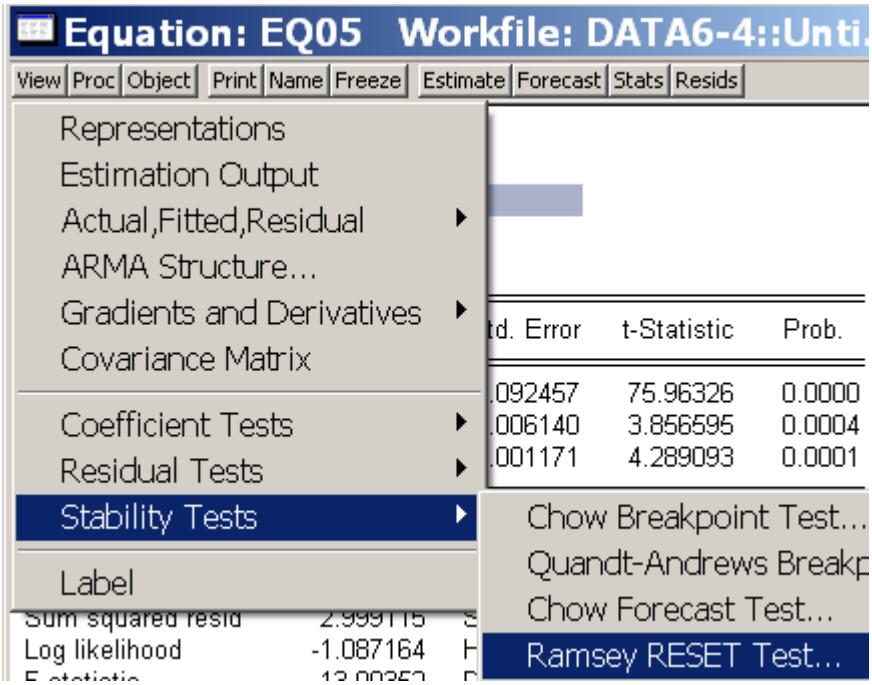
Ho: Mô hình không bị sai số đặc trưng (mô hình được xác định đúng)

H1: Mô hình bị sai số đặc trưng

P-value(F)=0.8827 (>0.05) nên ở độ tin cậy 95% ta chưa đủ cơ sở để bác bỏ Ho, hay có thể chấp nhận giả thuyết cho rằng mô hình được xác định đúng.

- Bên cạnh kiểm định RESET của Ramsey, bạn có thể phát hiện sai số đặc trưng bằng kiểm định tính phân phối chuẩn của phần dư, Durbin-Watson (khi có dữ liệu theo thời gian)…

Hình 2.1



Hình 2.2

Ramsey RESET Test:

F-statistic	0.1251	Prob. F(2,44)	**0.8827**
Log likelihood ratio	0.2779	Prob. Chi-Square(2)	0.8703

Test Equation:
Dependent Variable: LOG(WAGE)
Method: Least Squares
Date: 01/07/11 Time: 10:12
Sample: 1 49
Included observations: 49

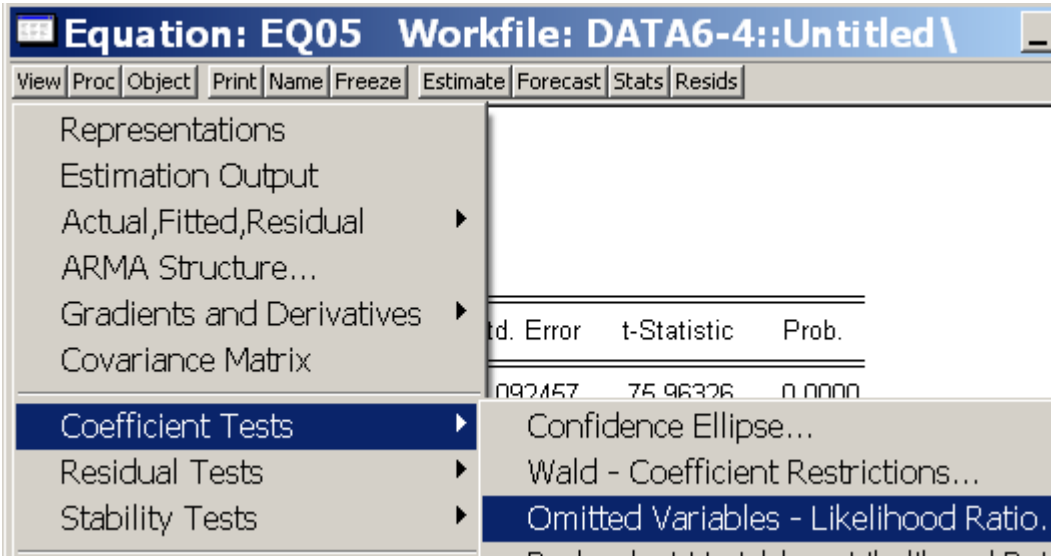
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-1164.0340	3046.5790	-0.3821	0.7042
EXPER	-6.1419	15.9710	-0.3846	0.7024
EDUC^2	-1.3029	3.3878	-0.3846	0.7024
FITTED^2	34.3385	89.6524	0.3830	0.7036
FITTED^3	-1.5085	3.9704	-0.3799	0.7058

R-squared	0.3648	Mean dependent var	7.4550
Adjusted R-squared	0.3070	S.D. dependent var	0.3127
S.E. of regression	0.2603	Akaike info criterion	0.2428
Sum squared resid	2.9822	Schwarz criterion	0.4358
Log likelihood	-0.9482	Hannan-Quinn criter.	0.3160
F-statistic	6.3170	Durbin-Watson stat	1.7309
Prob(F-statistic)	0.0004		

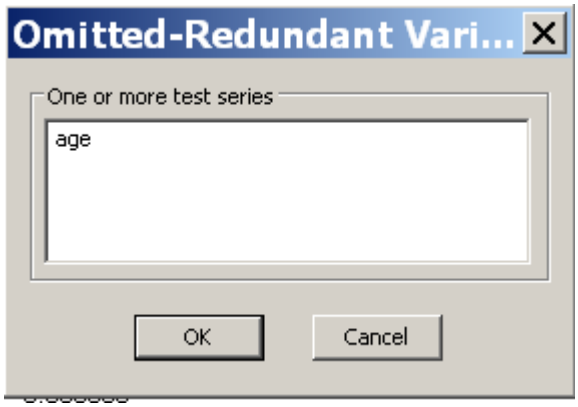
b. Anh Nam nghi ngờ rằng "mô hình 2 bị bỏ sót một biến quan trọng là biến age". Theo bạn, nghi ngờ này là đúng hay sai?

Tại cửa sổ Equation của phương trình 2, bạn hãy chọn **View\Coefficient test\Omitted variable – Likelihoot Ratio ;** sau đó nhập vào biến age → OK → Kết quả như Hình 2.5

Hình 2.3



Hình 2.4



Kết quả ở Hình 2.5 cho bạn kiểm định giả thuyết

Ho: Biến age không bị bỏ sót (β_{age} =0)

H1: Biến age bị bỏ sót

P-value(F)=0.824 (>0.05) nên chưa đủ cơ sở để bác bỏ Ho. Nói cách khác có thể chấp nhận giả thuyết cho rằng Mô hình 2 không bị bỏ sót biến age

🗝Bạn có thể thực hiện kiểm định nhân tử Lagrance để trả lời câu hỏi này…

Hình 2.5

Omitted Variables: AGE

F-statistic	0.0500	Prob. F(1,45)	**0.824**
Log likelihood ratio	0.0544	Prob. Chi-Square(1)	0.816

Test Equation:
Dependent Variable: LOG(WAGE)
Method: Least Squares
Date: 01/07/11 Time: 10:32
Sample: 1 49
Included observations: 49

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	6.990	0.174573	40.04283	0.000
EXPER	0.023	0.006692	3.454777	0.001
EDUC^2	0.005	0.001185	4.250427	0.000
AGE	0.001	0.003934	0.223646	**0.824**

R-squared	0.362	Mean dependent var	7.455
Adjusted R-squared	0.319	S.D. dependent var	0.313
S.E. of regression	0.258	Akaike info criterion	0.207
Sum squared resid	2.996	Schwarz criterion	0.361
Log likelihood	-1.060	Hannan-Quinn criter.	0.265
F-statistic	8.507	Durbin-Watson stat	1.722
Prob(F-statistic)	0.000		

c. Nếu mô hình 2 là mô hình đúng, nhưng bạn ước lượng mô hình sau:

$Ln(Wage)=\alpha_1+ \alpha_2 educ^2+u$ (3)

Trong trường hợp này, mô hình 3 sẽ bị sai số đặc trưng gì? Và hậu quả sẽ như thế nào?

- Mô hình bị thiếu biến quan trọng

- Do hệ số tương quan giữa exper và $educ^2$ = -0.282 và có ý nghĩa thống kê (P-value=0.049) nên $\hat{\alpha}_1$, và $\hat{\alpha}_2$ bị thiên lệch và không nhất quán, phương sai của hạng nhiễu bị ước lượng không đúng, SE($\hat{\alpha}_2$) bị ước lượng chệch, và khoảng tin cậy, cũng như những kiểm định thông thường (như kiểm định t, F) có thể đưa ra những kết luận sai.

Hình 2.6

Correlation t-Statistic Probability	EXPER	EDUC^2
EXPER	1.000000 ----- -----	
EDUC^2	-0.282081 -2.015710 0.0496	1.000000 ----- -----

d. Với mô hình 3, theo bạn, làm sao biết được biến exper có phải là biến bị bỏ sót hay không?

Hình 1.2 cho thấy biến exper có P-value = 0.0004 (<0.05) nên biến exper nên có ở trong mô hình. Nói cách khác, với mô hình 3, biến exper đã bị bỏ sót.

Bạn cũng có thể kiểm định điều này bằng thao tác tương tự như Bài 2b, hay thực hiện kiểm định nhân tử Lagrance … để trả lời câu hỏi này

Bài 3

Bạn hãy sử dụng dữ liệu VHLSS 2008, thực hiện các câu lệnh trên Stata đề trả lời các câu hỏi sau:

a. Trong mẫu khảo sát, hiện nay (tại thời điểm điều tra) có bao nhiêu người hiện đang đi học, đang nghỉ hè, bỏ học (m2ac5)?

. tab m2ac5

```
  5.HiÖn cã |
     ®i häc |      Freq.     Percent        Cum.
------------+-----------------------------------
         Cã |      4,531       11.84       11.84
    NghØ hÌ |      5,410       14.14       25.99
      Kh«ng |     28,312       74.01      100.00
------------+-----------------------------------
      Total |     38,253      100.00
```

. tab m2ac5

Trong 38253 người được khảo sát, tại thời điểm điều tra, có 4531 người hiện đang đi học, 5410 người đang nghỉ hè, và 28312 người không đi học

b. Với những người hiện đang đi học hoặc nghỉ hè, trung bình chi tiêu cho giáo dục (tạm thời bạn sử dụng biến m2ac13k để đo lường chi tiêu cho giáo dục của từng người) của nam là bao nhiêu? và của nữ là bao nhiêu?

```
. tab  m1ac2 if m2ac5<=2, sum (m2ac13k)

    2. Giíi | Summary of 13k.Tæng sè (a+b+...+i)
       tÝnh |        Mean   Std. Dev.       Freq.
------------+------------------------------------
        Nam |   1614.0948   2722.1407        5108
         N÷ |   1521.1268   2577.7371        4833
```

```
------------+------------------------------------
      Total |   1568.8967   2653.1921        9941
```

Trung bình chi tiêu cho giáo dục của nam là 1614.1 ngàn đ/năm, và của nữ là 1521.1 ngàn đ/năm

c. Bạn hãy tạo biến giả tên là gioi với 1 là nam, 0 là nữ; và thống kê xem trong số những người hiện đang đi học hoặc nghỉ hè thì có bao nhiêu nam, bao nhiêu nữ.

. tab m1ac2 if m2ac5<=2

```
    2. Giíi |
       tÝnh |      Freq.     Percent        Cum.
------------+-----------------------------------
        Nam |      5,108       51.38       51.38
         N÷ |      4,833       48.62      100.00
------------+-----------------------------------
      Total |      9,941      100.00
```

Trong số những người hiện đang đi học hoặc nghỉ hè, có 5108 nam và 4833 nữ

d. Bạn hãy ước lượng hàm hồi quy tuyến tính thể hiện ảnh hưởng của tuổi, giới, đến chi tiêu cho giáo dục (Chỉ ước lượng hàm hồi quy trên những người mà tại thời điểm điều tra đang đi học hoặc nghỉ hè)

. reg chigd tuoi gioi if m2ac5<=2

```
      Source |       SS       df       MS              Number of obs =    9941
-------------+------------------------------           F(  2,  9938) = 1130.28
       Model |  1.2967e+10     2  6.4834e+09           Prob > F      =  0.0000
    Residual |  5.7005e+10  9938  5736081.11           R-squared     =  0.1853
-------------+------------------------------           Adj R-squared =  0.1851
       Total |  6.9972e+10  9940  7039428.34           Root MSE      =    2395

------------------------------------------------------------------------------
       chigd |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        tuoi |   203.2432   4.278268    47.51   0.000     194.8569    211.6295
       gioi |   68.67071   48.06328     1.43   0.153    -25.54307    162.8845
       _cons |  -1045.171   64.07087   -16.31   0.000    -1170.763   -919.5791
------------------------------------------------------------------------------
```

Phương trình hồi quy là:

Chigd=-1045.17 +203.24tuoi+68.67gioi+ $\hat{u}$

e. Bạn hãy ước lượng hàm hồi quy tuyến tính (có sử dụng trọng số) thể hiện ảnh hưởng của tuổi, tuổi bình phương, giới, đến chi tiêu cho giáo dục (Chỉ ước lượng hàm hồi quy trên những người mà tại thời điểm điều tra đang đi học hoặc nghỉ hè)

```
. reg chigd tuoi tuoibp gioi [pw= hhszwt] if  m2ac5<=2
(sum of wgt is   1.0913e+08)
Linear regression                                      Number of obs =    9941
                                                       F(  3,  9937) =  227.50
                                                       Prob > F      =  0.0000
                                                       R-squared     =  0.1514
                                                       Root MSE      =  2588.7
------------------------------------------------------------------------------
             |               Robust
       chigd |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
```

```
-------------+----------------------------------------------------------------
        tuoi |   144.8307   26.88878     5.39   0.000     92.12323    197.5381
      tuoibp |   1.632666   .9951162     1.64   0.101    -.3179639    3.583295
        gioi |   52.34552   66.75542     0.78   0.433    -78.50863    183.1997
       _cons |  -523.1452   163.7174    -3.20   0.001    -844.0646   -202.2259
------------------------------------------------------------------------------
```

f. Bạn hãy viết một do-file thực hiện các công việc trên, các câu lệnh trong do-file này là gì?

Hình 3.1

Do-file Editor - baitap10f.do

File Edit Tools View

baitap10f.do | Untitled.do

```
1   /*       Tap lam 1 Do-file don gian          */
2   set mem 300m
3   use "C:\VHLSS2008\Data\Hhold\muc123a.dta", clear
4   tab m2ac5
5   tab  m1ac2 if m2ac5<=2, sum (m2ac13k)
6   gen gioi= m1ac2
7   recode gioi 2=0
8   tab  m1ac2 if m2ac5<=2
9   gen chigd= m2ac13k
10  gen tuoi= m1ac5
11  reg chigd tuoi gioi if  m2ac5<=2
12
13  /*luu lai file muc123a.dta vao thu muc khac*/
14  save "C:\VHLSS2008\tam\muc123a.dta", replace /*option replace giup luu de len file co*/
15
16  *Mo file hhexpe08.dta, giu lai nhung bien can thiet, sap xep
17  use "C:\VHLSS2008\Data\Hhold\hhexpe08.dta", clear
18  keep tinh huyen xa diaban hoso wt9 hhszwt urban08 reg8
19  sort tinh huyen xa diaban hoso
20  save "C:\VHLSS2008\tam\hhexpe08adj.dta",replace
21
22  *Mo lai file muc123a.dta da luu tru luc truoc
23  use "C:\VHLSS2008\tam\muc123a.dta", clear
24  sort tinh huyen xa diaban hoso
25  merge m:1  tinh huyen xa diaban hoso using "C:\VHLSS2008\tam\hhexpe08adj.dta"
26  drop _merge
27
28  gen tuoibp=m1ac5^2
29  reg chigd tuoi tuoibp gioi [pw= hhszwt] if  m2ac5<=2
30  save "C:\VHLSS2008\tam\muc123a_hhexpe08adj.dta", replace
31
```

g. Trong mẫu điều tra, theo bạn, có bao nhiêu người hiện nay vừa đang đi học vừa có việc làm? Với những người vừa được hỏi ở Mục 123a và Mục 4a, có bao nhiêu người ở thành thị, bao nhiêu người ở nông thôn? Bạn hãy cho biết các câu lệnh trong do-file để thực hiện điều này là gì?

Từ câu f, bạn đã tạo được file dữ liệu mới **muc123a_hhexpe08adj.dta.** file này chứa các biến của muc123a, một số biến của file hhexpe08.dta (trọng số cá nhân, trọng số hộ, urban08, reg8

Để trả lời câu g. Bạn cần nối các file **muc4a.dta** vào file **muc123a_hhexpe08adj.dta**

Bạn có thể thực hiện các câu lệnh sau tiếp theo với các câu lệnh ở Do-file trên

Hình 3.2

Do-file Editor - bt10_cau3.do*

File Edit Tools View

bt10_cau3.do* Untitled.do

```
27
28    gen tuoibp=m1ac5^2
29    reg chigd tuoi tuoibp gioi [pw= hhszwt] if  m2ac5<=2
30    save "C:\VHLSS2008\tam\muc123a_hhexpe08adj.dta", replace
31
32    *Cau 3g - Baitap10
33    use "C:\VHLSS2008\Data\Hhold\muc4a.dta", clear
34    sort tinh huyen xa diaban hoso matv
35    save "C:\VHLSS2008\tam\muc4a_sorted.dta", replace
36
37    use "C:\VHLSS2008\tam\muc123a_hhexpe08adj.dta", clear
38    sort tinh huyen xa diaban hoso matv
39    merge 1:1  tinh huyen xa diaban hoso matv using "C:\VHLSS2008\tam\muc4a_sorted.dta"
40
41    tab _merge
42    keep if _merge==3
43    drop _merge
44    count
45
46    tab  m2ac5 m4ac2
47    tab urban08
48    save "C:\VHLSS2008\tam\muc123a_hhexpe08adj_muc4a.dta", replace
49
50
```

```
. tab  m2ac5 m4ac2

 5.HiÖn cã |    2. Cã viÖc lµm
    ®i häc |        Cã      Kh«ng |     Total
-----------+----------------------+----------
        Cã |       409      3,552 |     3,961
   NghØ hÌ |       482      4,576 |     5,058
     Kh«ng |    21,689      4,446 |    26,135
-----------+----------------------+----------
     Total |    22,580     12,574 |    35,154
```

Số người vừa đang đi học (bao gồm cả những người trả lời hiện nay đang nghỉ hè), vừa đang đi làm là 409+482= 891 người

```
. tab urban08

    Khu vùc |      Freq.     Percent        Cum.
------------+-----------------------------------
      Urban |      8,853       25.18       25.18
          2 |     26,301       74.82      100.00
------------+-----------------------------------
      Total |     35,154      100.00
```

Với những người vừa được hỏi ở Mục 123a và vừa được hỏi ở Mục 4a, có 8853 người ở thành thị, 26301 người ở nông thôn

🗝Ghi chú thêm

Tuy nhiên, khi muốn ước lượng tỷ lệ người ở thành thị, ở nông thôn được đúng hơn cho tổng hể, với VHLSS bạn cần quan tâm đến trọng số, trong trường hợp này, mỗi dòng phân tích là một cá nhân (một thành viên), vì vậy bạn có thể sử dụng biến hhszwt

```
. tab urban08 [aw=hhszwt]

    Khu vùc |      Freq.     Percent        Cum.
------------+-----------------------------------
      Urban |9,620.89707       27.37       27.37
          2 | 25,533.103       72.63      100.00
------------+-----------------------------------
      Total |     35,154      100.00
```

Tỷ lệ người dân sống ở thành thị là 27.37%, tỷ lệ người dân sống ở nông thôn là 72.63%

Trong file dữ liệu, khi mỗi dòng là một hộ, và bạn muốn ước lượng cho tổng thể về chỉ tiêu nào đó của hộ (ước lượng tỷ lệ, trung bình, hay thực hiện hàm hồi quy …) cần sử dụng trọng số hộ là biến wt9 (trọng số hộ)