



Phân tích dữ liệu bằng phần mềm SPSS 12.0



PHÂN TÍCH DỮ LIỆU BẰNG PHẦN MỀM SPSS 12.0*

PHẦN 1

Các nội dung chính trong phần này:

- 1. Khởi động SPSS**
- 2. Giới thiệu một số menu chính trong cửa sổ SPSS**
- 3. Mở dữ liệu từ tập tin của SPSS và của phần mềm EXCEL**
- 4. Nhập dữ liệu trực tiếp từ bàn phím và bằng Copy và Paste**
- 5. Vẽ đồ thị scatter**
- 6. Thống kê mô tả dữ liệu**
- 7. Kiểm tra sự tương quan**
- 8. Thêm biến và thêm quan sát**

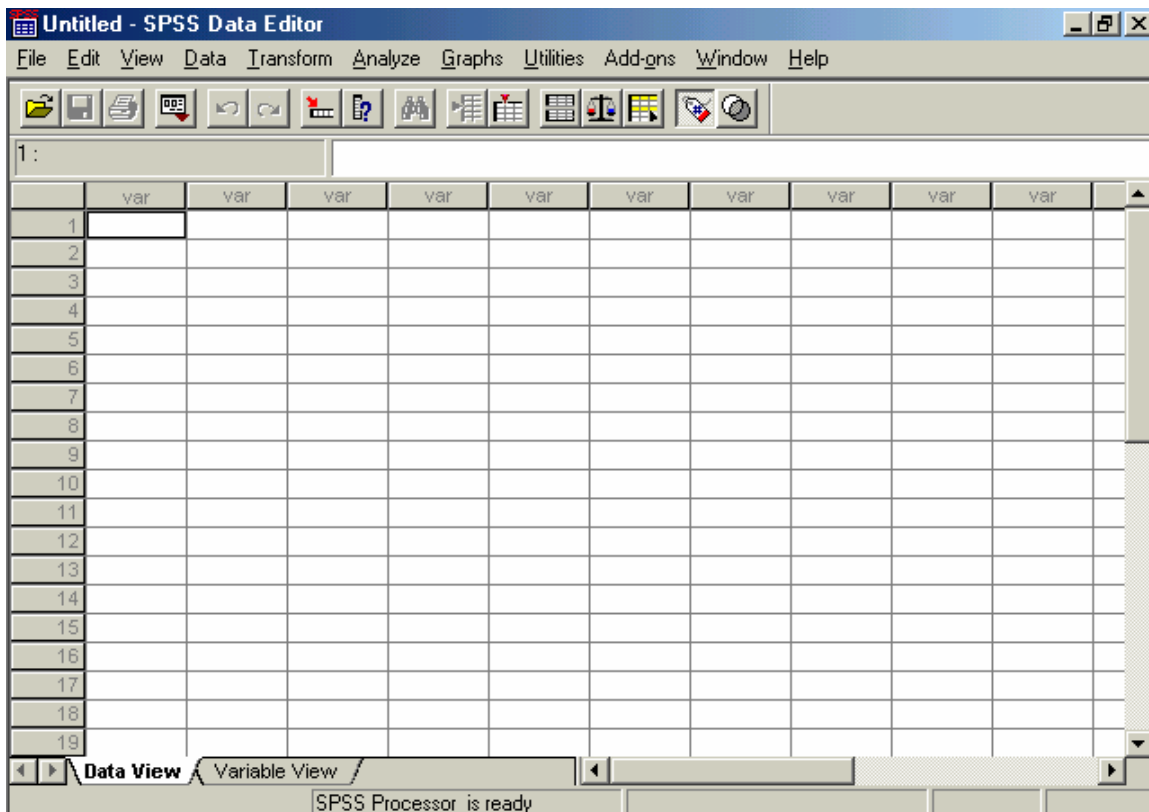
* SPSS là một nhãn hiệu đã đăng ký bản quyền của SPSS Inc.

1. Khởi động SPSS

- Cách 1: Nhấp đúp vào biểu tượng SPSS trên màn hình.
- Cách 2: Vào Start, Program, SPSS for Windows



HÌNH 1



2.

File Edit View Data Transform Analyze Graphs Utilities Add-ons Windows Help

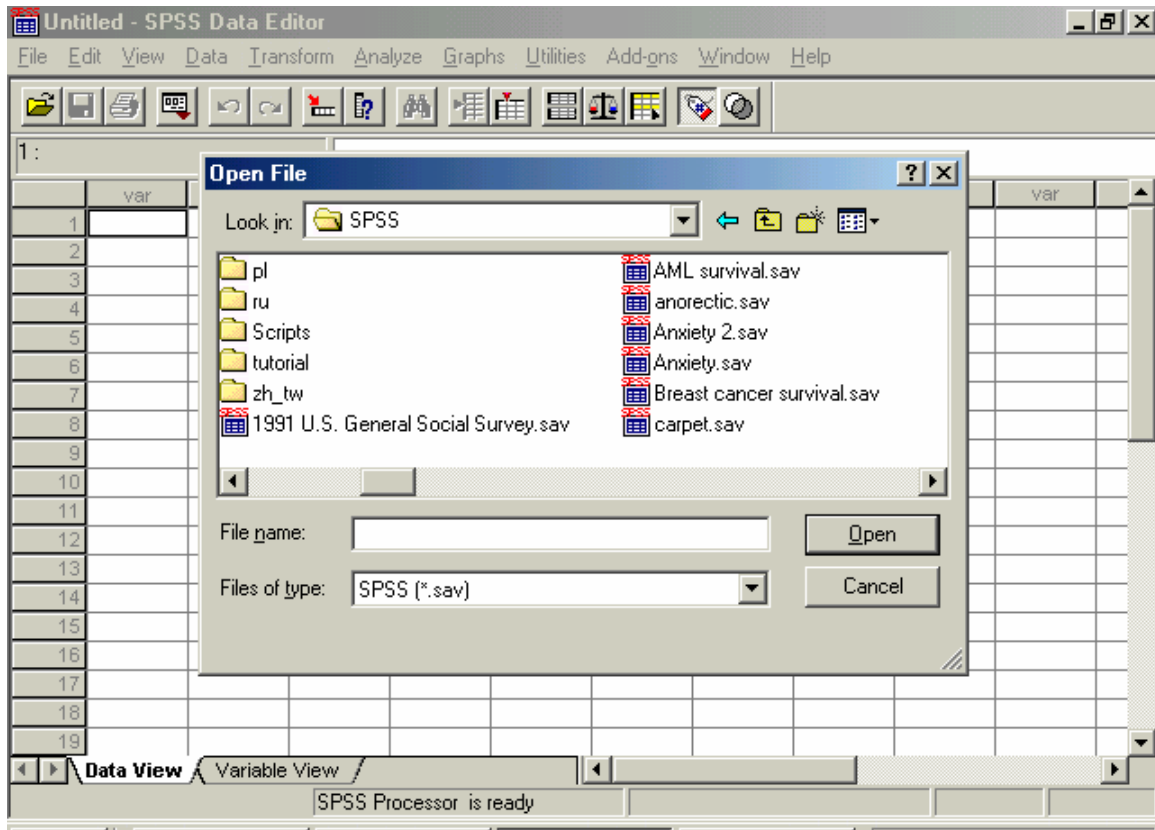
- File:** Chứa nội dung về: Tạo mới dữ liệu, Mở dữ liệu từ các tập tin của SPSS và của các phần mềm khác, Lưu trữ dữ liệu, In ấn và Thoát khỏi SPSS.
- Edit:** Chứa nội dung về: Thực hiện lại hoặc hủy bỏ một thao tác vừa mới thực hiện, Sao chép và cắt dán dữ liệu, Chỉnh sửa một số nội dung tùy chọn chung (Options).
- Analyze:** Chứa nội dung về: Thống kê mô tả, Lập bảng biểu, chạy các loại hồi quy.
- Graph:** Chứa nội dung về: Vẽ đồ thị.
- Help:** Chứa nội dung về: Những trợ giúp cho người sử dụng.

3. Mở dữ liệu từ tập tin của SPSS và của phần mềm EXCEL

a. Từ tập tin của SPSS

Vào Menu **File, Open, Data**. Sau đó chọn loại tập tin có phần mở rộng ***.sav**

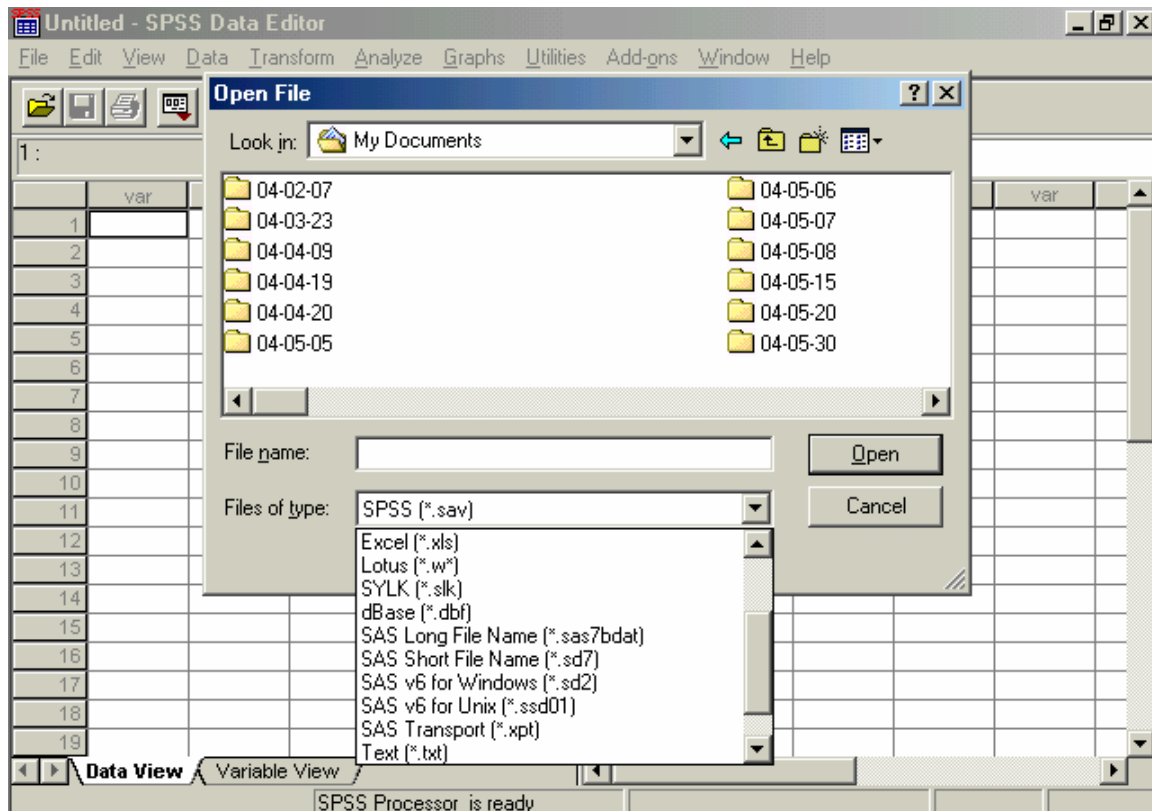
HÌNH 2



b. Từ tập tin của phần mềm EXCEL

Vào Menu **File, Open, Data**. Sau đó, vào mục **Files of type** để chọn loại tập tin cần truy xuất dữ liệu. Ở đây, chúng ta quan tâm đến tập tin của **EXCEL**.

HÌNH 3



Ví dụ:

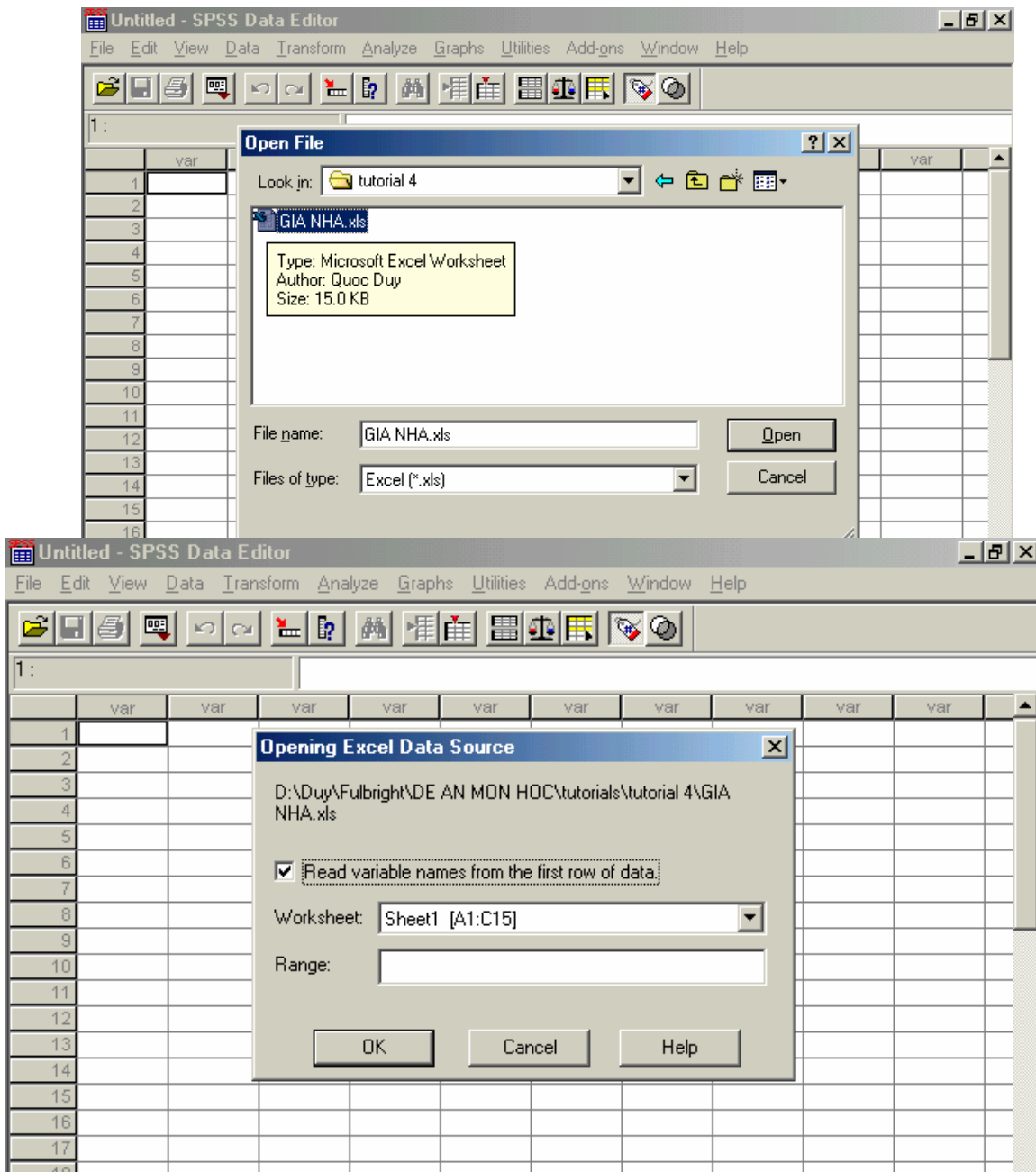
Có 1 tập tin EXCEL chứa dữ liệu về giá nhà và diện tích như đã học. Nội dung của tập tin này bắt đầu từ A1 đến C15. Bây giờ chuyển nội dung của tập tin này sang SPSS.

HÌNH 4

	A	B	C	D	E	F	G	H	I
1	STT	PRICE	SQFT						
2	1	199.9	1065						
3	2	288	1254						
4	3	235	1300						
5	4	285	1577						
6	5	239	1600						
7	6	293	1750						
8	7	285	1800						
9	8	365	1870						
10	9	295	1935						
11	10	290	1948						

Trước tiên, cần đóng tập tin này lại (nếu như đang mở ra xem nó trong EXCEL). Sau đó, vào SPSS, chọn **File, Open, Data**. Sau đó chọn **Files of type** và chọn **Excel (*.xls)**. Sau đó chọn tập tin Excel cần sử dụng.

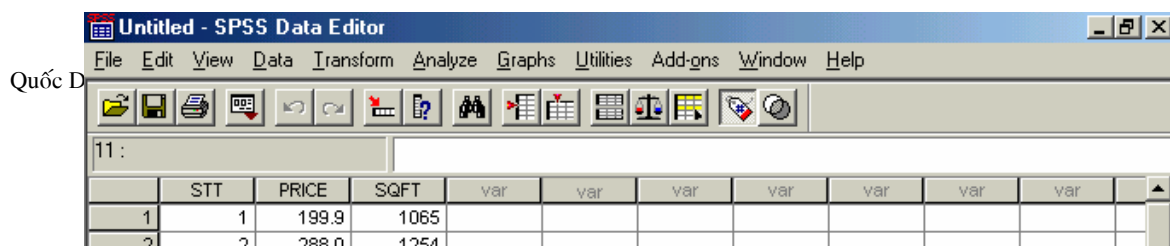
HÌNH 5



Dòng **Worksheet** cho biết là trong tập tin **GIA NHA.XLS** có bao nhiêu Sheet chứa dữ liệu và mức độ dàn trải của dữ liệu. Nếu chọn dòng chữ **Read variable names from the first row of data**, thì SPSS sẽ lấy tên biến giống như tên đã được nhập sẵn trong Excel.

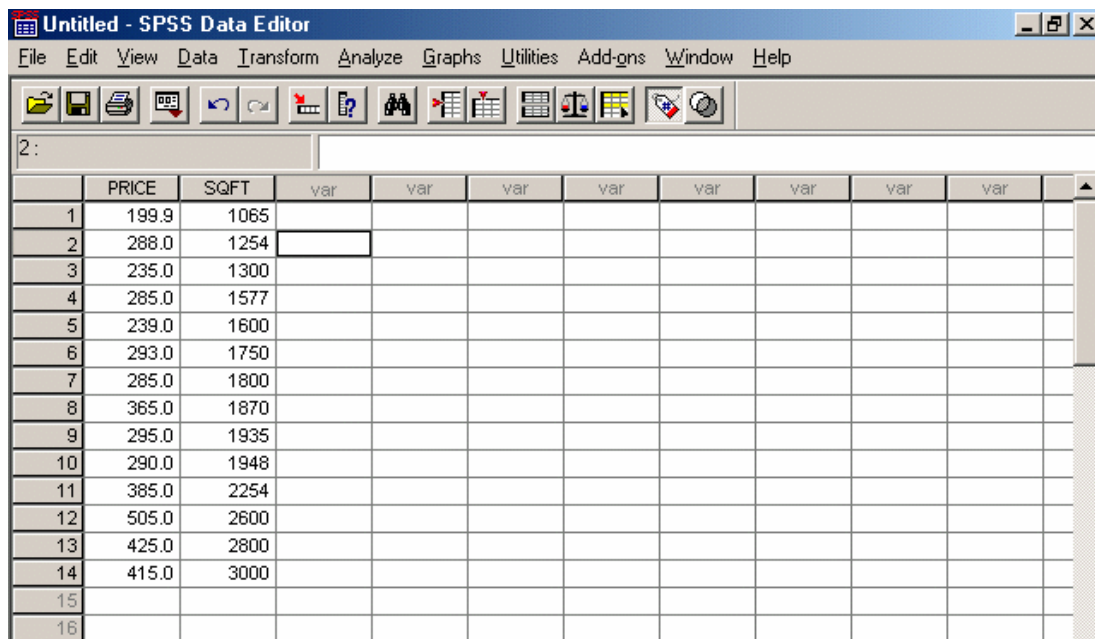
- Nếu chúng ta lấy hết toàn bộ nội dung của file Excel thì nhấn **OK**.

HÌNH 7



- Nếu chúng ta chỉ lấy một phần nội dung của file Excel thì nhập thông tin vào **Range**. Chẳng hạn, nếu chúng ta không lấy cột STT trong file Excel thì gõ vào **Range: B1:C15** rồi nhấn **OK**. Tức là chúng ta chỉ lấy cột PRICE và SQFT.

HÌNH 8



The screenshot shows the SPSS Data Editor window titled "Untitled - SPSS Data Editor". The menu bar includes File, Edit, View, Data, Transform, Analyze, Graphs, Utilities, Add-ons, Window, and Help. The toolbar contains various icons for file operations and data manipulation. The data grid shows a table with the following data:

	PRICE	SQFT	var	var	var	var	var	var	var	var	var
1	199.9	1065									
2	288.0	1254									
3	235.0	1300									
4	285.0	1577									
5	239.0	1600									
6	293.0	1750									
7	285.0	1800									
8	365.0	1870									
9	295.0	1935									
10	290.0	1948									
11	385.0	2254									
12	505.0	2600									
13	425.0	2800									
14	415.0	3000									
15											
16											

4. Nhập dữ liệu trực tiếp từ bàn phím và bằng Copy và Paste

Từ tập tin trong Excel, chọn dòng và cột dữ liệu cần sao chép rồi **Copy**.

HÌNH 9

The screenshot shows a Microsoft Excel window titled 'Microsoft Excel - GIA NHA.xls'. The active cell is B1, containing the formula '= PRICE'. The data table is as follows:

	A	B	C	D	E	F	G	H	I
1	STT	PRICE	SQFT						
2	1	199.9	1065						
3	2	228	1254						
4	3	235	1300						
5	4	285	1577						
6	5	239	1600						
7	6	293	1750						
8	7	285	1800						
9	8	365	1870						
10	9	295	1935						
11	10	290	1948						
12	11	385	2254						

Sau đó, mở SPSS, vào **File, New, Data**. Sau đó chọn vị trí và Paste dữ liệu vừa Copy từ Excel vào.

HÌNH 10

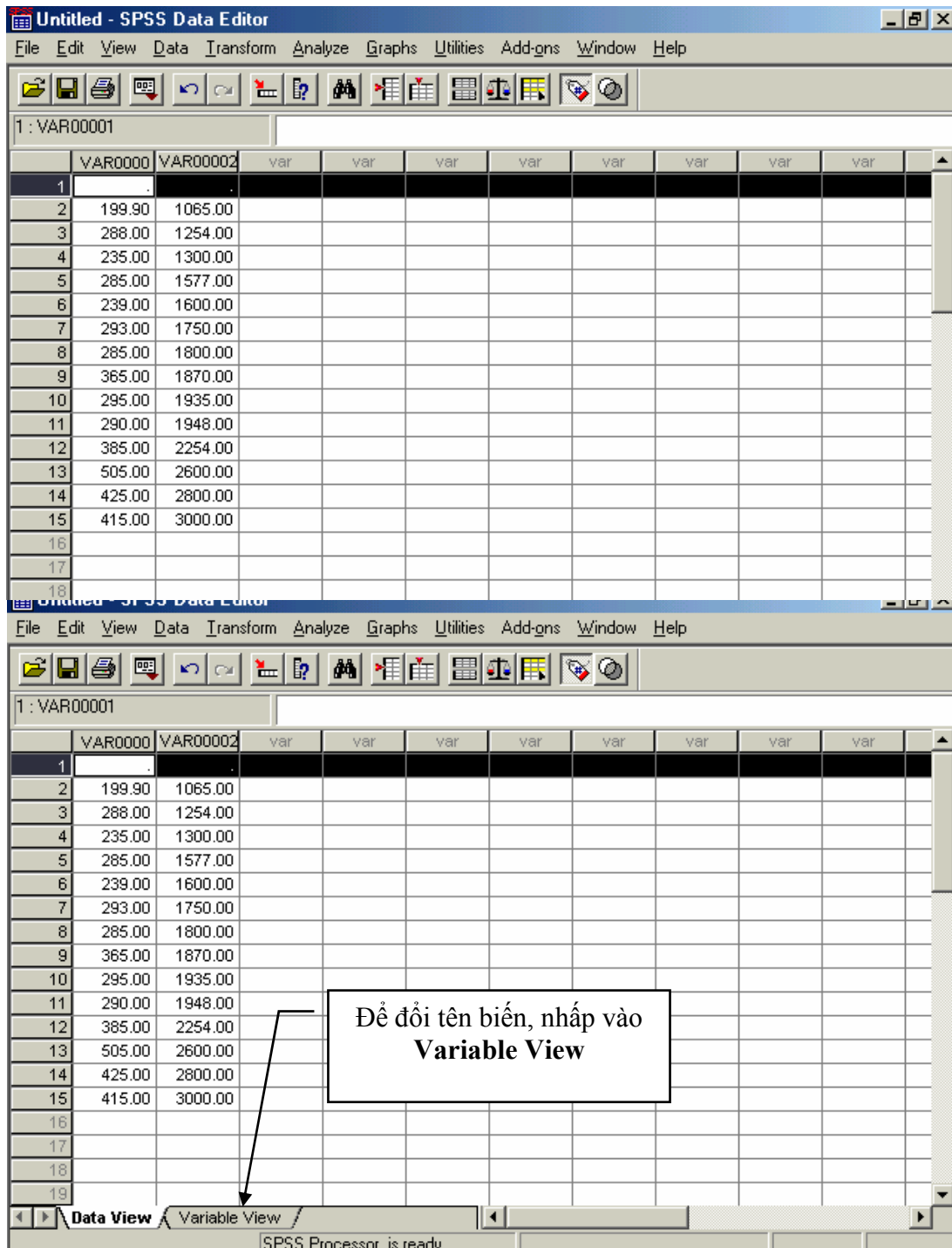
The screenshot shows the SPSS Data Editor window titled 'Untitled - SPSS Data Editor'. The data table is as follows:

	VAR00001	VAR00002	var	var	var	var	var	var	var	var
1	.	.								
2	199.90	1065.00								
3	288.00	1254.00								
4	235.00	1300.00								
5	285.00	1577.00								
6	239.00	1600.00								
7	293.00	1750.00								
8	285.00	1800.00								
9	365.00	1870.00								
10	295.00	1935.00								
11	290.00	1948.00								
12	385.00	2254.00								
13	505.00	2600.00								
14	425.00	2800.00								
15	415.00	3000.00								
16										

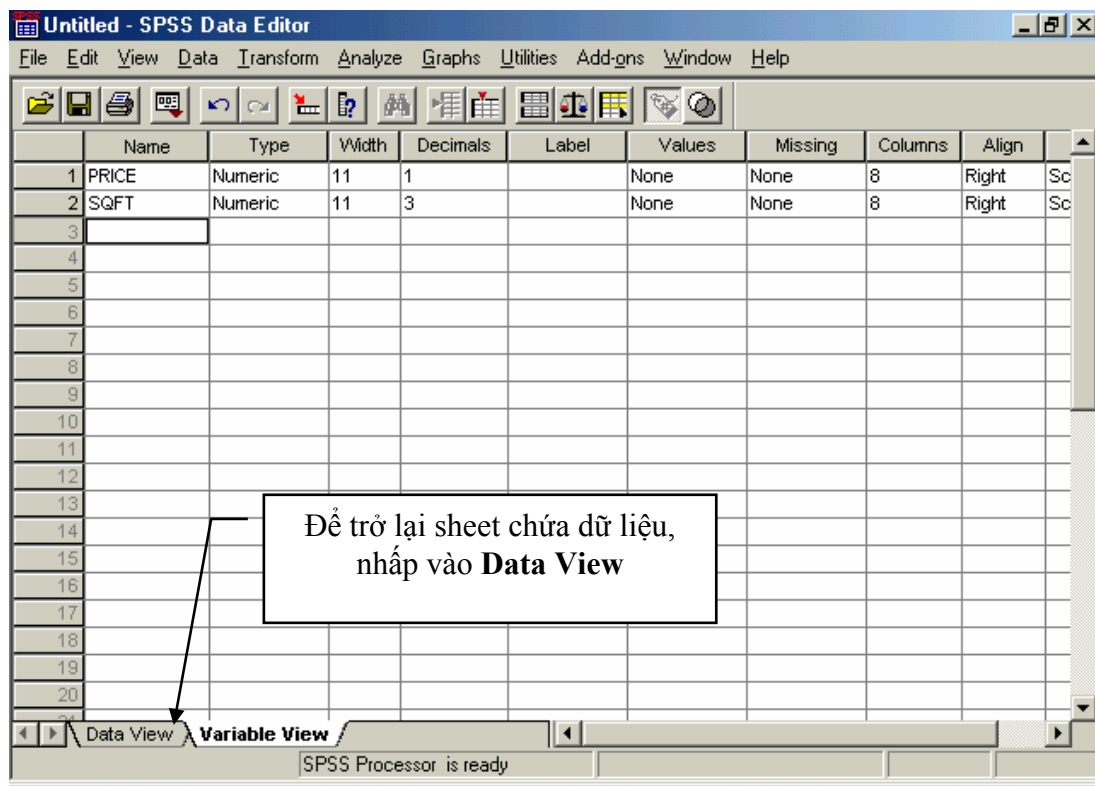
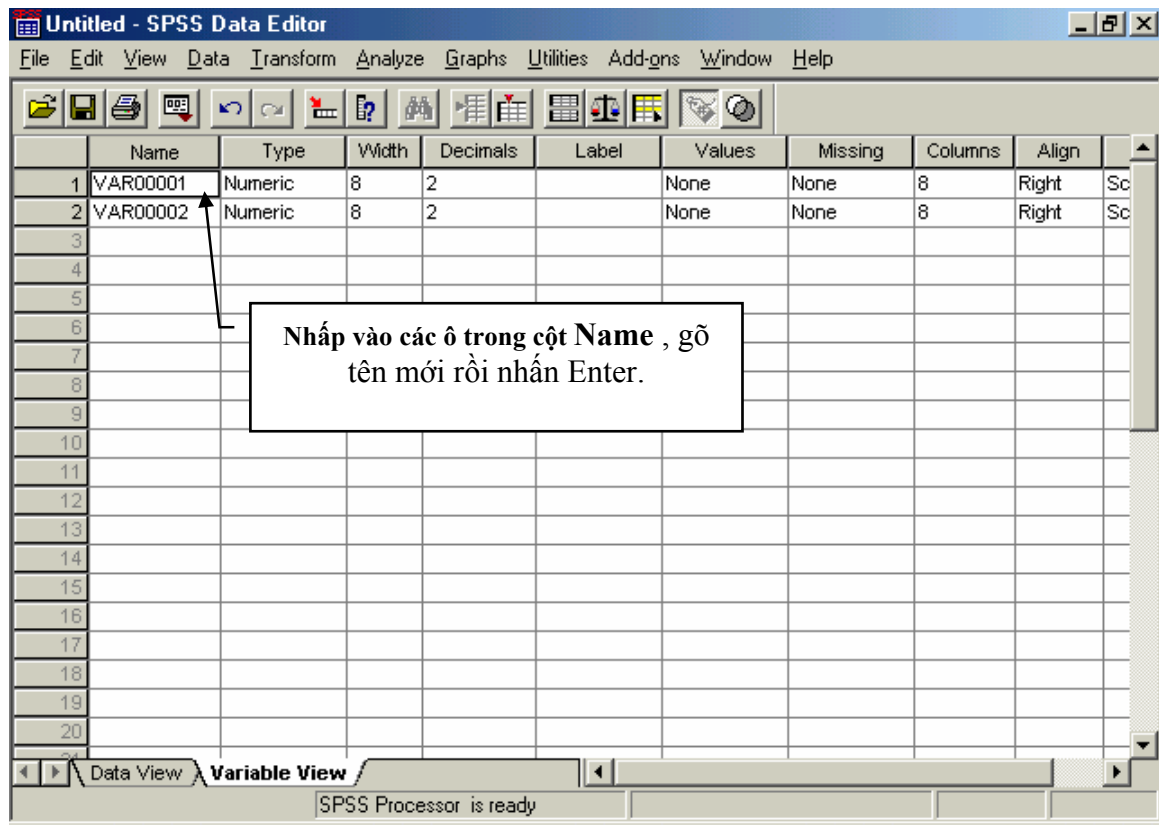
Do Copy trực tiếp nên dòng đầu tiên chứa tên biến, không phải là giá trị nên SPSS bỏ trống dòng đó. Mặc khác, tên biến được mặc định là **VAR0001** và **VAR0002**.

Để bỏ đi dòng đầu tiên, chúng ta nhấp vào dòng số 1, sau đó click chuột phải và chọn **Clear**.

HÌNH 11



HÌNH 13



5. Vẽ đồ thị scatter

Vào Menu **Graphs**, chọn **Scatter**

HÌNH 15

Ch n Simple

Sau ã nh p Define

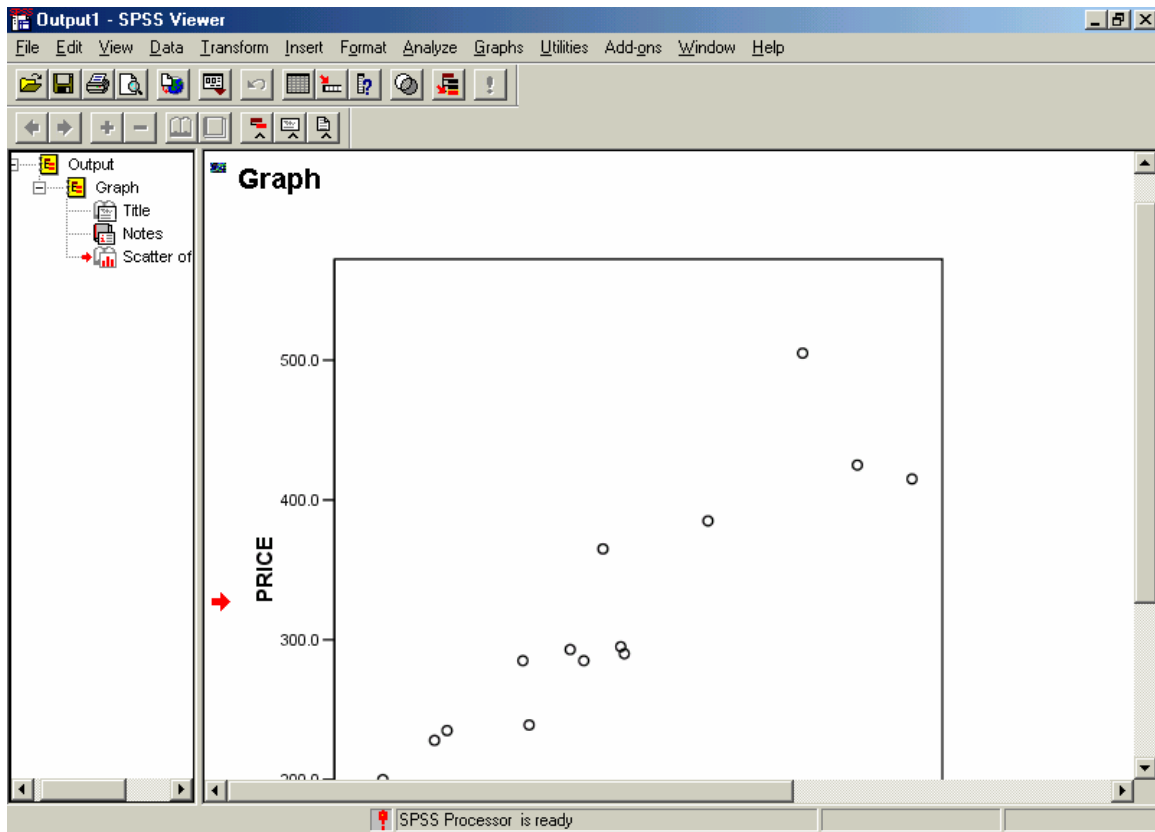
Nhập vào biến **PRICE** là biến phụ thuộc.

Sau đó nhập vào nút này để đưa biến **PRICE** vào trục tung

STT	PRICE	SQFT	var	var	var
1	199.9	1065			
2	228.0	1254			
3	235.0	1200			
4	285.0				
5	235.0				
6	293.0				
7	285.0				
8	365.0				
9	295.0				
10	290.0	1948			
11	385.0	2254			
12	505.0	2600			
13	425.0	2800			
14	415.0	3000			
15					
16					

Sau khi nhập **OK**, kết quả vẽ đồ thị sẽ hiện ra trong một cửa sổ mới có tên là **Output1**. Và kết quả của những lệnh khác cũng sẽ thể hiện trong cửa sổ này.

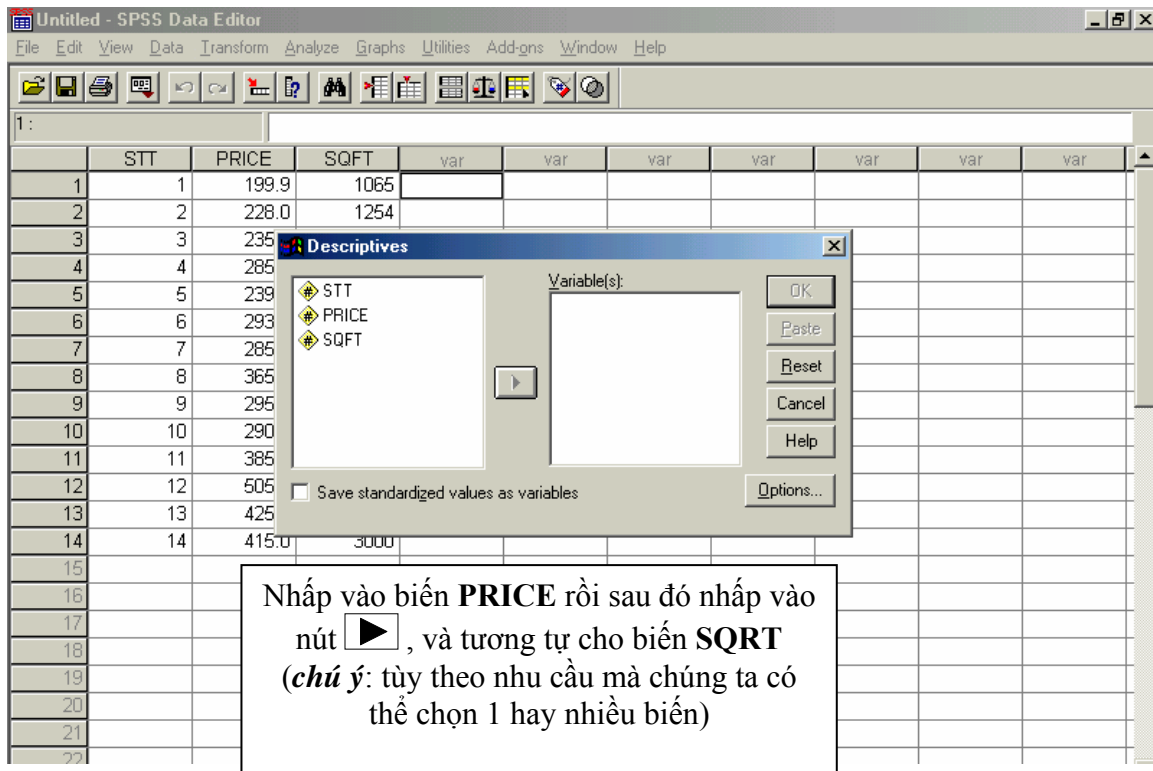
HÌNH 18



6. Thống kê mô tả dữ liệu

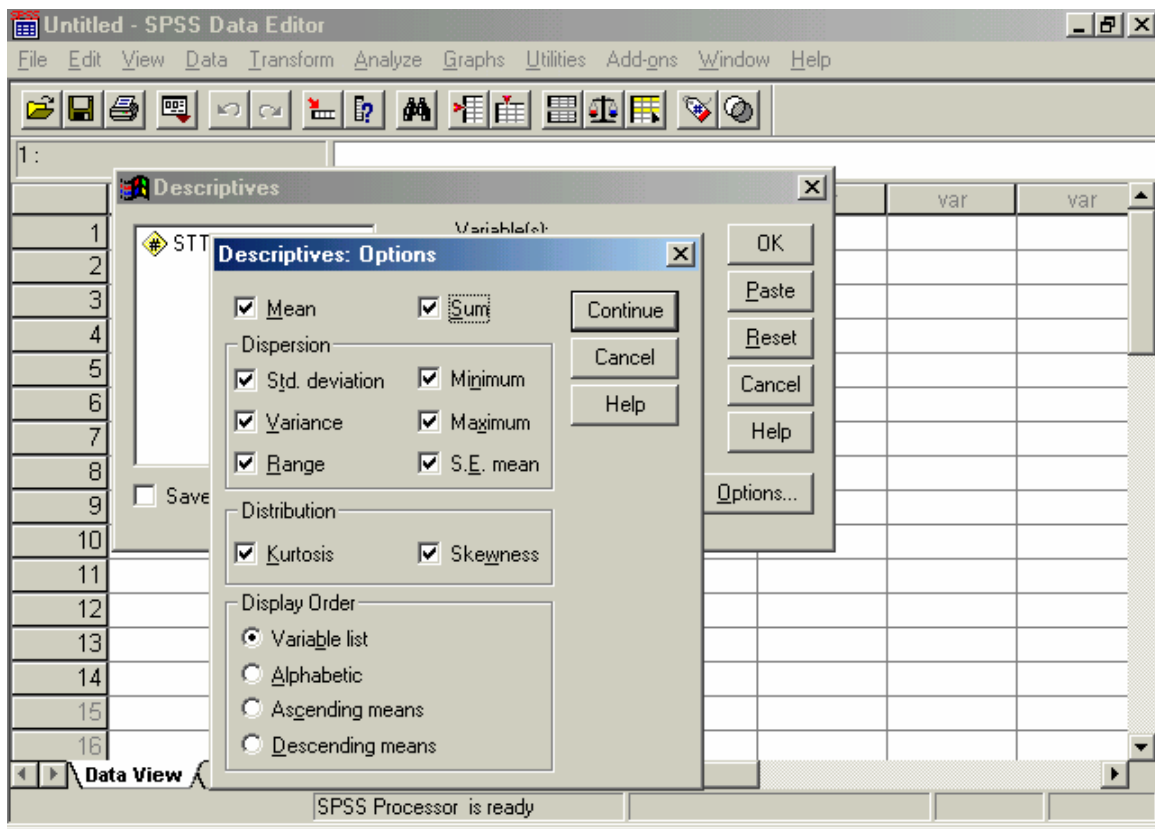
Vào Menu **Analyze, Descriptive Statistics, Descriptives...**

HÌNH 19



Để lựa chọn các chỉ tiêu cần quan tâm, nhấp vào **Options**. Nếu muốn chọn chỉ tiêu nào thì nhấp vào ô trắng bên cạnh chỉ tiêu đó.

HÌNH 20



Sau khi chọn xong các chỉ tiêu cần xem xét, nhấp vào nút **Continue** để trở hộp thoại **Descriptives** rồi bấm **OK**. Kết quả thống kê mô tả sẽ hiện ra tại cửa sổ **Output1**, bảng kết quả được bố trí nằm ngang.

7. Thêm bớt biến và quan sát

a. Thêm quan sát

- **Nói tiếp:** Quan sát muốn thêm vào là nối tiếp theo các quan sát trước thì chúng ta nhấp vào ô cần thêm giá trị và gõ vào bình thường giống như Excel.

HÌNH 21

15: PRICE 232

	STT	PRICE	SQFT	var	var	var	var	var
1	1	199.9	1065					
2	2	228.0	1254					
3	3	235.0	1300					
4	4	285.0	1577					
5	5	239.0	1600					
6	6	293.0	1750					
7	7	285.0	1800					
8	8	365.0	1870					
9	9	295.0	1935					
10	10	290.0	1948					
11	11	385.0	2254					
12	12	505.0	2600					
13	13	425.0	2800					
14	14	415.0	3000					
15		232						
16								

- **Chèn:** Quan sát muốn thêm vào là xen giữa các quan sát đã có thì chúng ta sẽ **chọn dòng cần chèn** rồi **click chuột phải**, chọn **Insert Cases**.

HÌNH 22

12: STT 12

	STT	PRICE	SQFT	var	var	var	var	var
1	1	199.9	1065					
2	2	228.0	1254					
3	3	235.0	1300					
4	4	285.0	1577					
5	5	239.0	1600					
6	6	293.0	1750					
7	7	285.0	1800					
8	8	365.0	1870					
9	9	295.0	1935					
10	10	290.0	1948					
11	11	385.0	2254					
12	12	505.0	2600					
		425.0	2800					
		415.0	3000					

giống như Excel.

HÌNH 23

The screenshot shows the SPSS Data Editor window titled 'Untitled - SPSS Data Editor'. The menu bar includes File, Edit, View, Data, Transform, Analyze, Graphs, Utilities, Add-ons, Window, and Help. The toolbar contains various icons for file operations and data manipulation. The main window displays a data table with the following structure:

	STT	PRICE	SQFT	var	var	var	var	var
2	2	228.0	1254					
3	3	235.0	1300					
4	4	285.0	1577					
5	5	239.0	1600					
6	6	293.0	1750					
7	7	285.0	1800					
8	8	365.0	1870					
9	9	295.0	1935					
10	10	290.0	1948					
11	11	385.0	2254					
12	.	456	.					
13	12	505.0	2600					
14	13	425.0	2800					
15	14	415.0	3000					
16								
17								

The status bar at the bottom indicates 'SPSS Processor is ready'.

b. Thêm biến

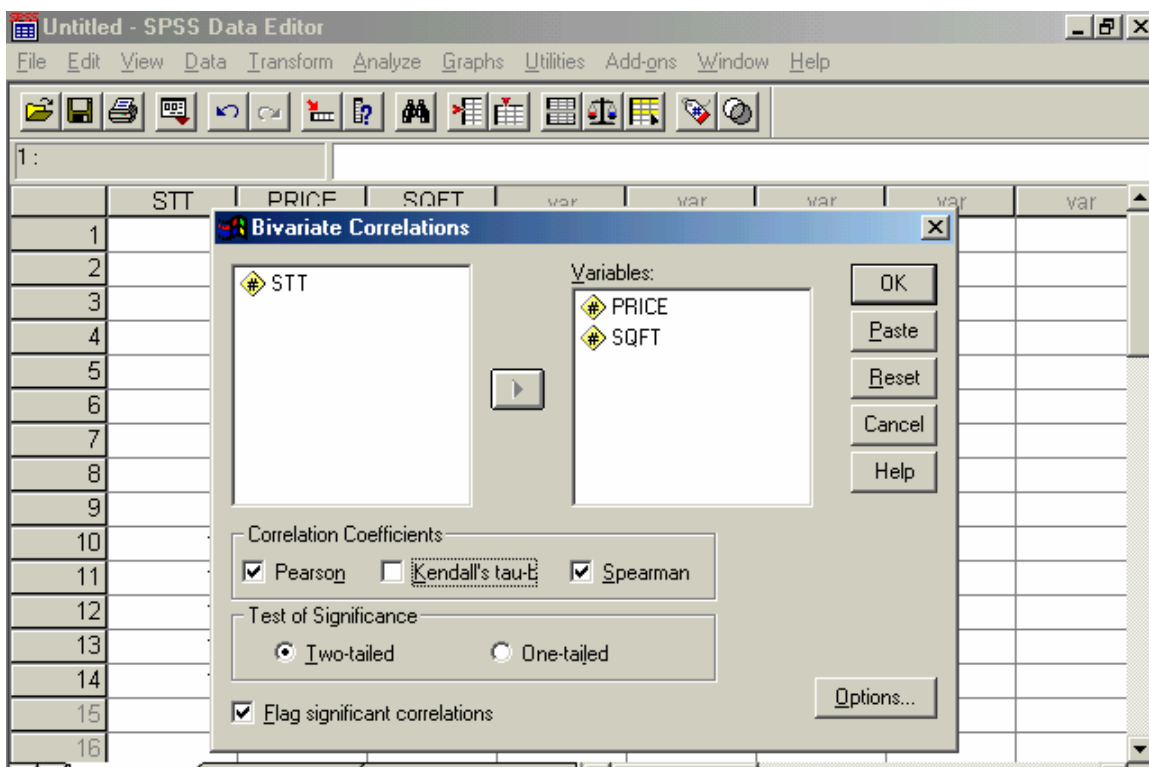
Thao tác để thêm biến cũng giống như trong Excel, chúng ta có thể gõ trực tiếp vào trong các ô của biến mới. Sau đó vào **Variable View** để đặt tên cho biến mới đó.

8. Kiểm tra sự tương quan

Vào Menu **Analyze**, chọn **Correlate, Bivariate...**

Sau khi đưa hai biến cần xem xét tương quan vào ô **Variables**, chúng ta có thể tùy chọn các loại hệ số tương quan. Trong trường hợp này chúng ta thử chọn **Pearson** và **Spearman**.

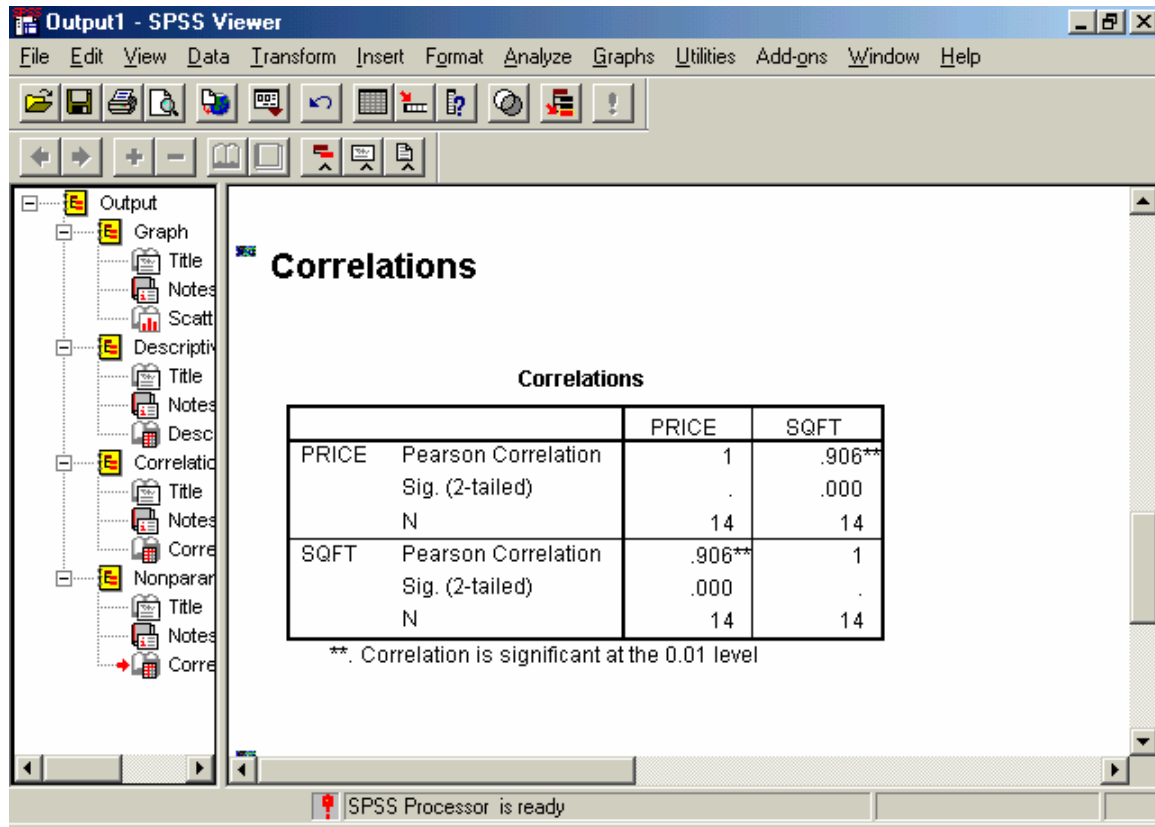
HÌNH 24



Sau khi bấm **OK** , kết quả sẽ được hiện ra trong cửa sổ **Output 1**.

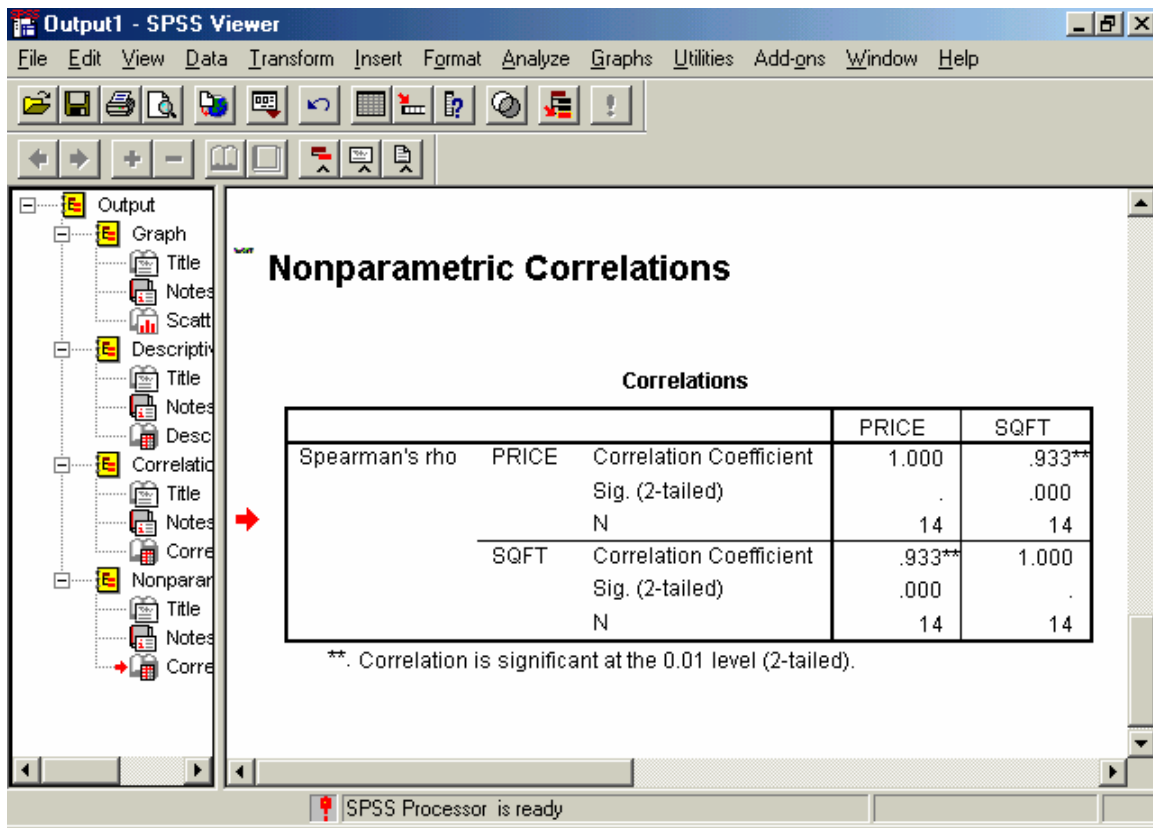
Đây là bảng kết quả của hệ số **Pearson**.

HÌNH 25



Còn đây là kết quả của hệ số Spearman.

HÌNH 26



Trong cả 2 bảng trên, những giá trị có ý nghĩa thống kê được đánh dấu bằng **

PHÂN TÍCH DỮ LIỆU BẰNG PHẦN MỀM SPSS 12.0*

PHẦN 2

Các nội dung chính trong phần này:

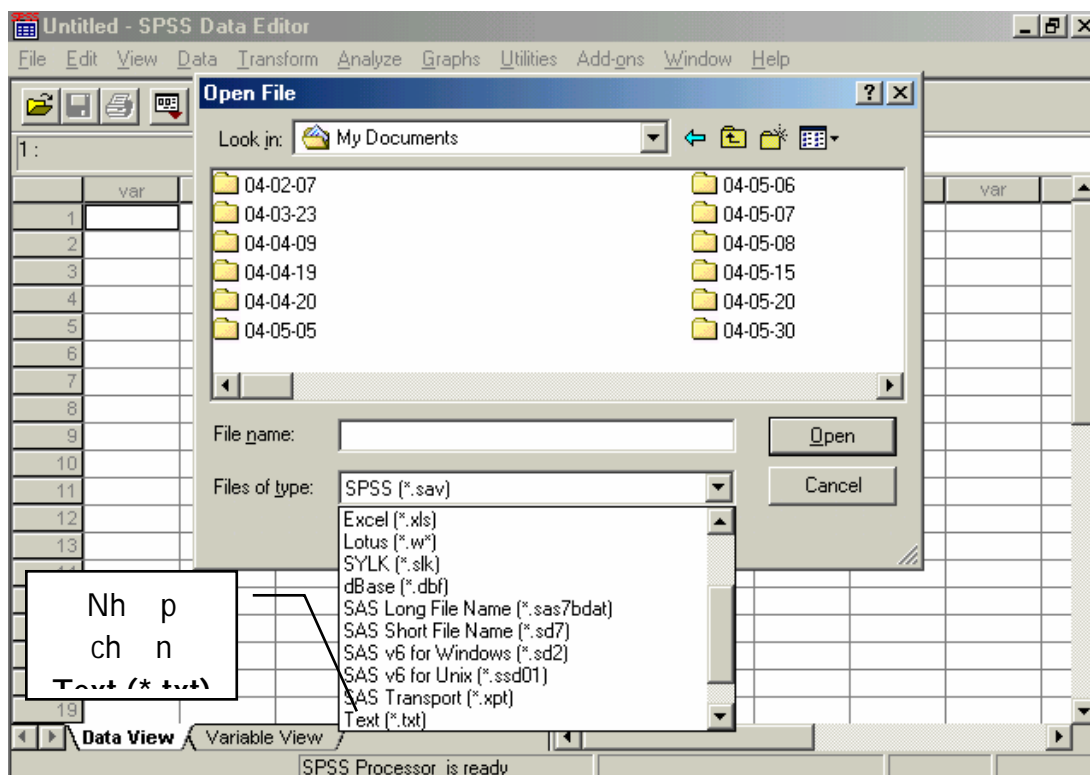
1. Mở dữ liệu từ tập tin dạng text
2. Xem xét sự tương quan giữa các biến
3. Hồi quy OLS (trường hợp đơn biến)
4. Hồi quy OLS (trường hợp đa biến)
5. Hồi quy trong trường hợp có hiện tượng đa cộng tuyến hoàn hảo

* SPSS 12.0 là sản phẩm đã đăng ký của SPSS Inc.

1. Mở dữ liệu từ tập tin dạng text

Vào Menu **File, Open, Data**. Sau đó, vào mục **Files of type** để chọn loại tập tin cần truy xuất dữ liệu. Ở đây, chúng ta quan tâm đến tập tin dạng text (*.txt).

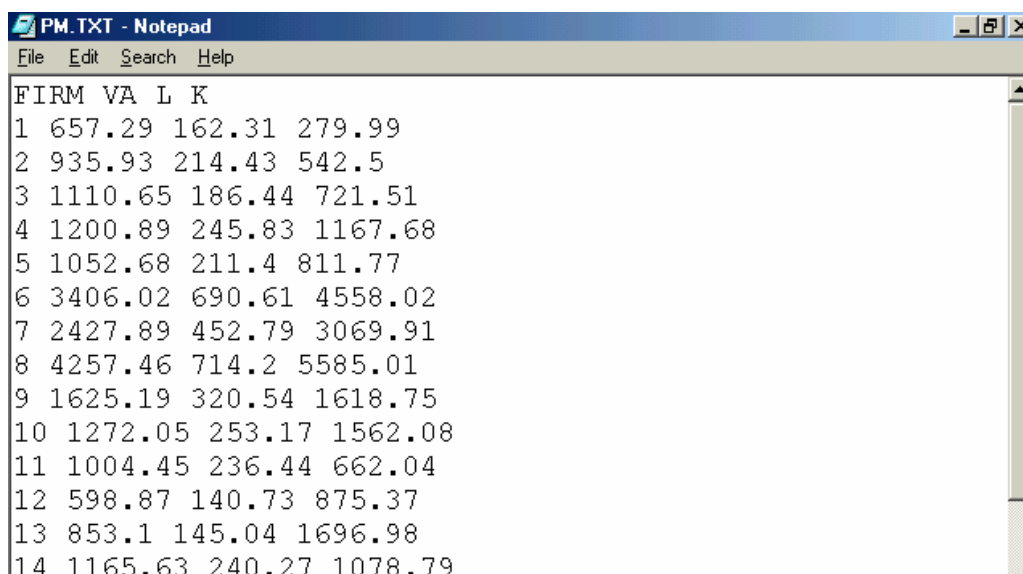
HÌNH 1



Ví dụ:

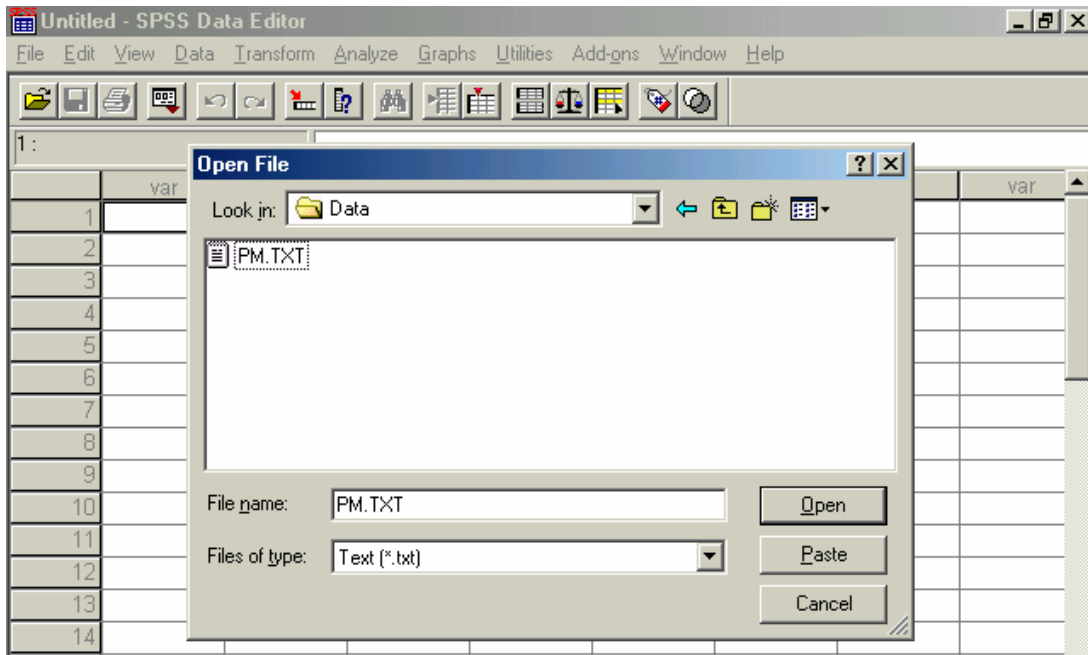
Chúng ta sẽ mở tập tin **pm.txt** chứa dữ liệu về **FIRM, VA, K** và **L** đã thực hành trong **EVIEWS**.

HÌNH 2



Trước tiên, cần đóng tập tin này lại (nếu như đang mở ra xem). Sau đó, vào **SPSS**, chọn **File, Open, Data**. Vào **Files of type** và chọn **Text (*.txt)**. Sau đó chọn tập tin text cần sử dụng là **pm.txt**

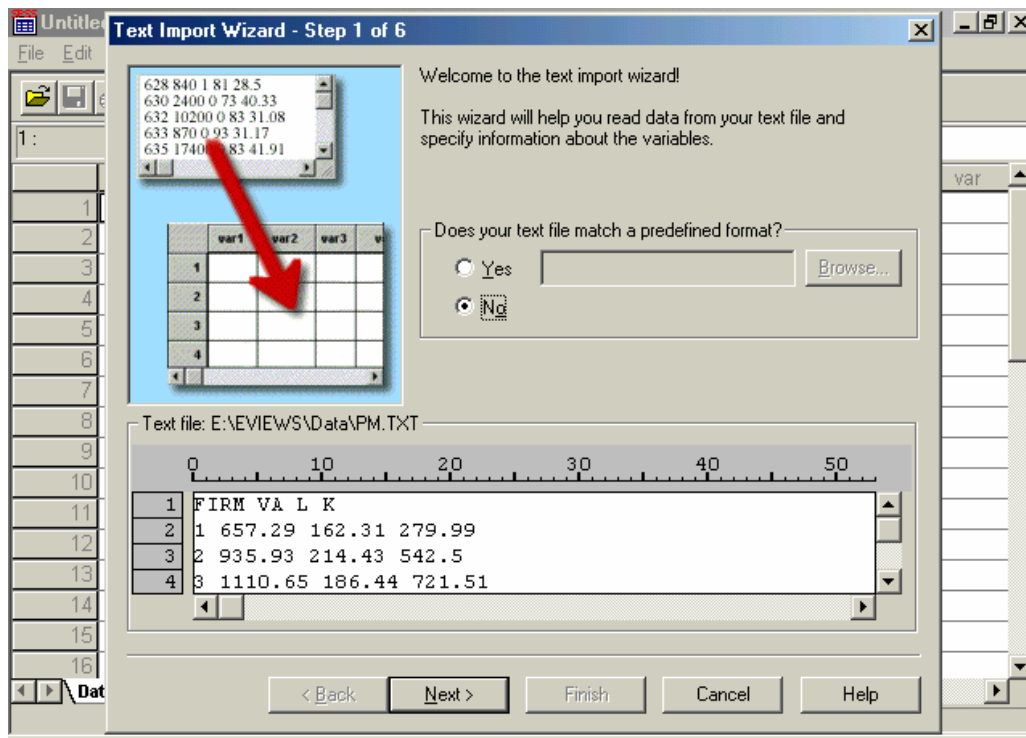
HÌNH 3



Việc nhập nội dung từ tập tin Text trải qua 6 bước.

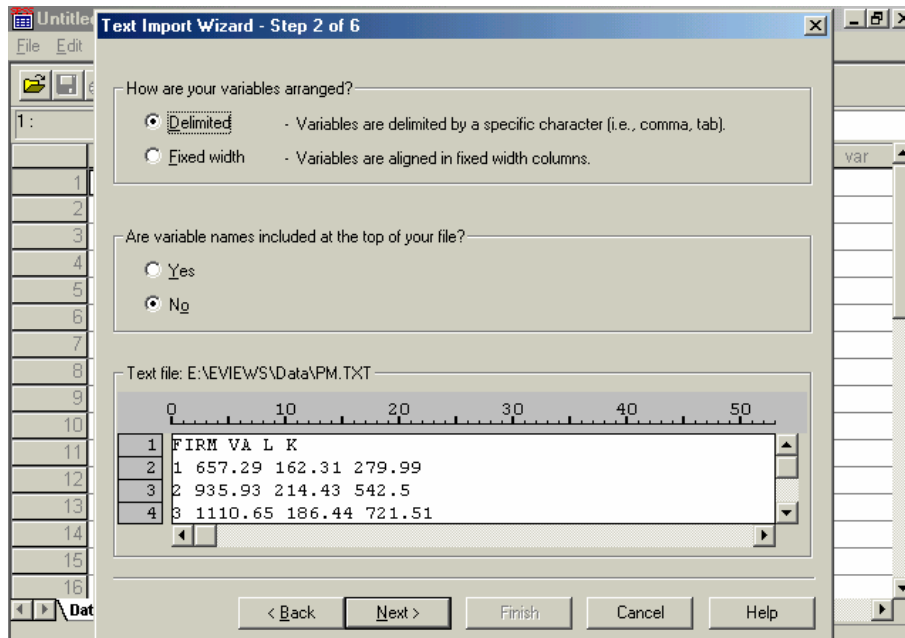
- **Bước 1 (Hình 4):** Trong Bước 1, SPSS mặc định sẵn ở chế độ **No**, chúng ta chỉ việc nhấp **Next**.

HÌNH 4



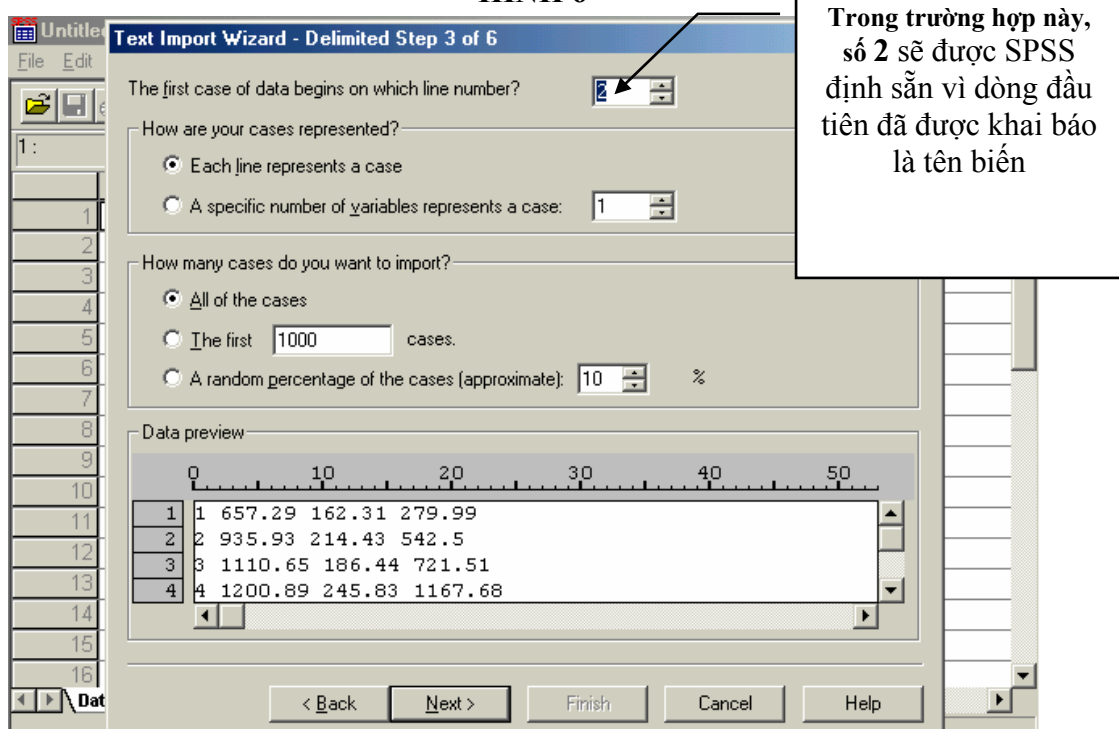
- **Bước 2 (Hình 5):** Trong Bước 2, cần chú ý dòng chữ **Are variable names included at the top of your file?**, ngụ ý hỏi có phải tên biến nằm ở dòng đầu tiên của tập tin đó không? Nếu phải thì nhấp **Yes**, ngược lại thì **No**. Trong trường hợp của tập tin **pm.txt** thì dòng đầu tiên có chứa tên biến nên chúng ta nhấp vào **Yes**. Sau đó, tiếp tục nhấp vào nút **Next**.

HÌNH 5



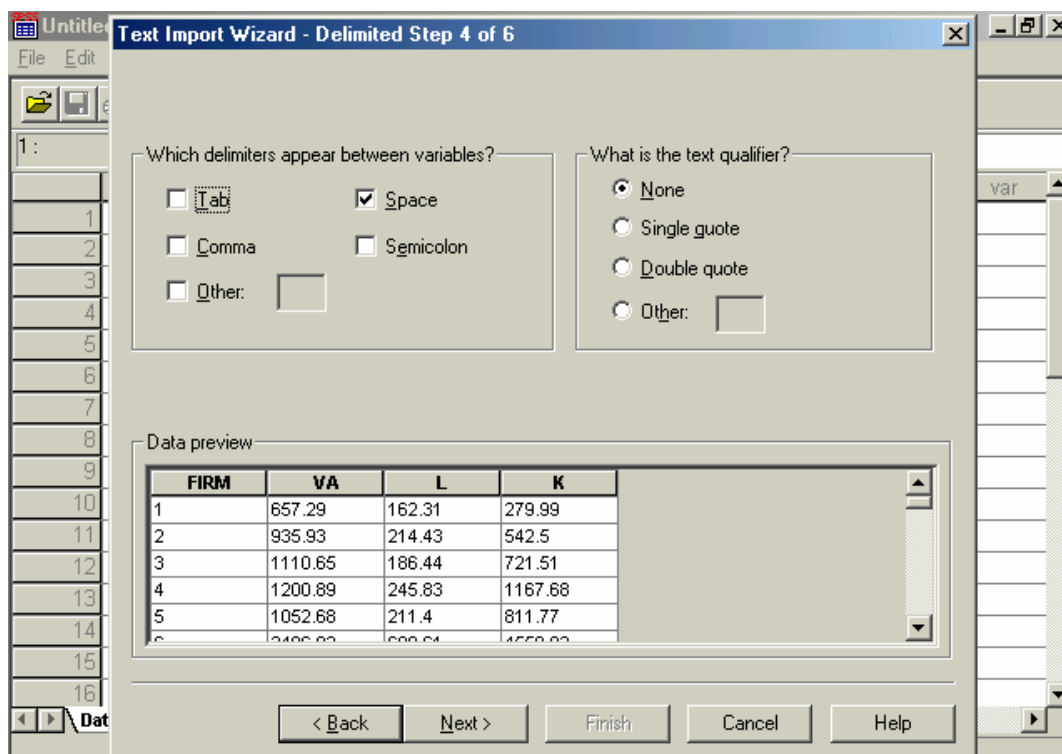
- **Bước 3 (Hình 6):** Chú ý dòng chữ **The first case of data begins on which line number?**, ngụ ý hỏi dữ liệu sẽ bắt đầu từ dòng thứ mấy. Khai báo xong, bấm **Next**.

HÌNH 6



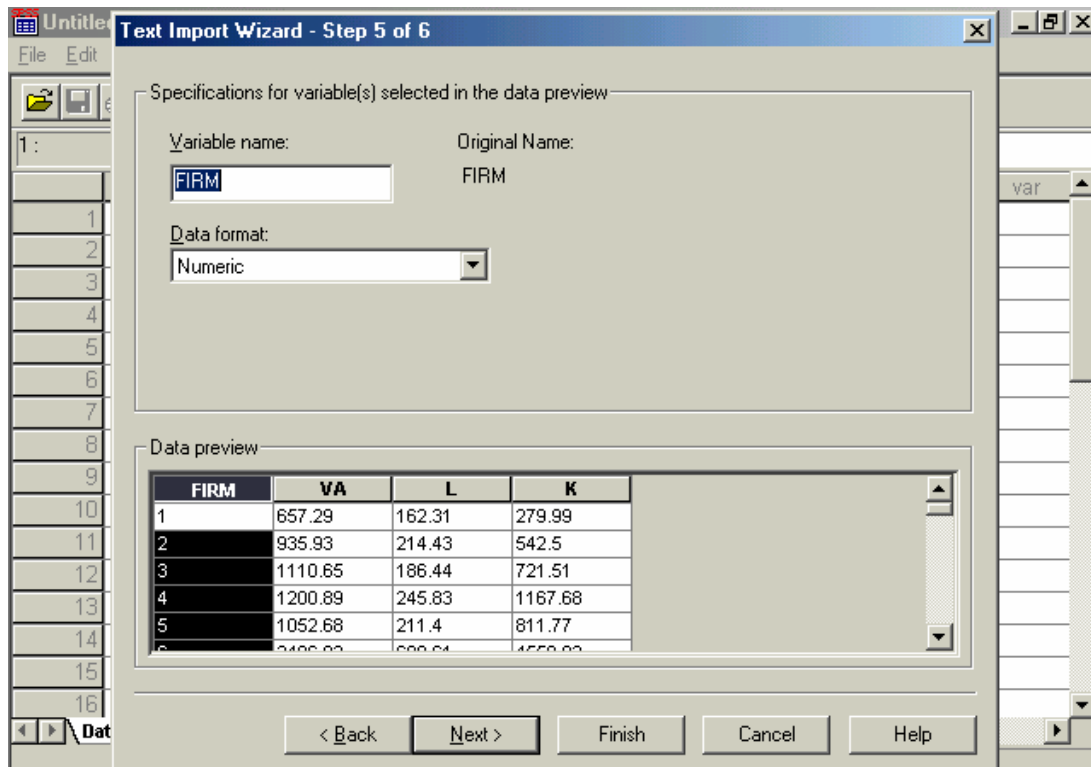
- **Bước 4 (Hình 7):** Chú ý đến câu **Which delimiters appear between variables?**, ngụ ý hỏi dữ liệu của các biến được phân định bằng cách nào. Bằng **Tab**, bằng khoảng trống (**Space**), bằng dấu phẩy (**Comma**), dấu chấm phẩy (**Semicolon**) hay ở dạng cụ thể nào khác (**Other**). Ở đây SPSS thường mặc định tại vị trí **Space** và trong trường hợp của tập tin **pm.txt**, dữ liệu giữa các biến đã được phân cách bằng khoảng trống nên chúng ta không cần phải chọn gì thêm mà chỉ việc nhấn **Next**.

HÌNH 7



- **Bước 5 (Hình 8):** Bước 8 là 1 bước mà SPSS cần chúng ta xác nhận lại xem tên biến và định dạng dữ liệu mà SPSS đã nhận diện là đúng chưa. Nếu chúng ta muốn thay đổi tên biến khác với tên gốc của nó thì sẽ gõ tên mới vào hộp thoại phía dưới dòng chữ **Variable name**, hay muốn định dạng dữ liệu lại thì sẽ vào **Data format** để điều chỉnh. Còn nếu thấy không cần phải thay đổi gì thì tiếp tục nhấn **Next** hoặc nhấn **Finish**.

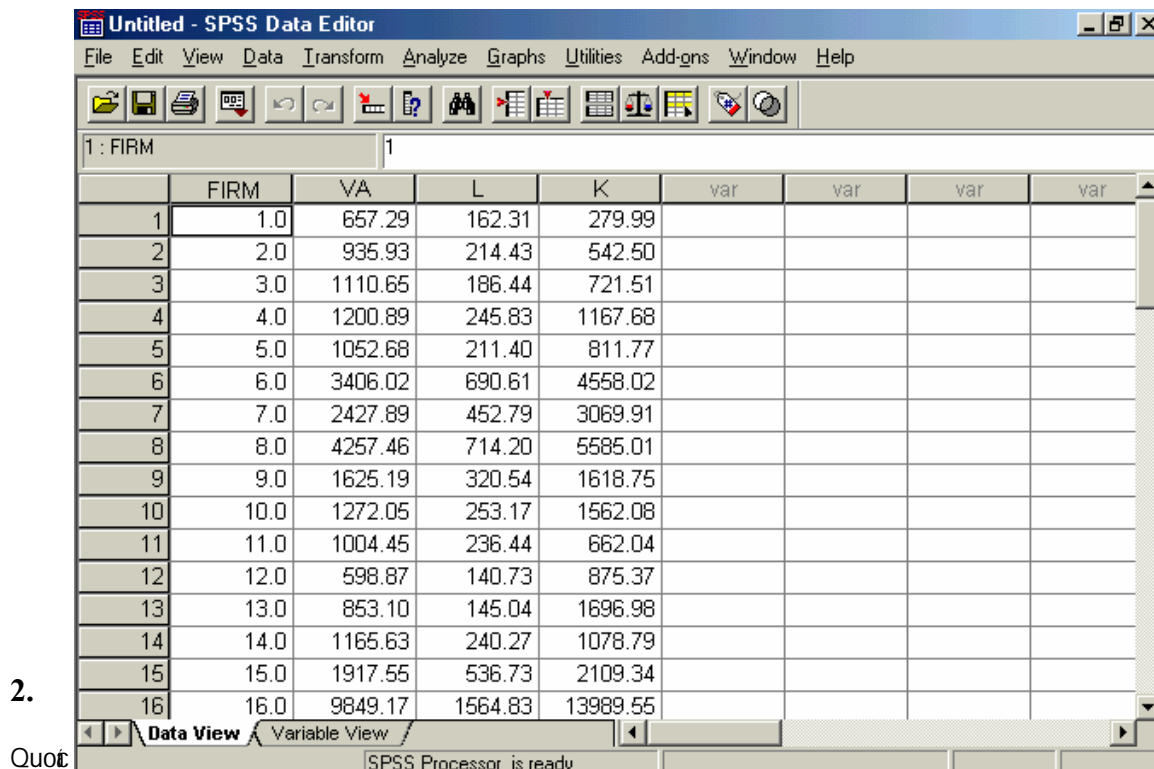
HÌNH 8



- **Bước 6:** Bước 6 chỉ là 1 bước thủ tục và chúng ta không cần để ý đến nội dung mà chỉ cần nhấn **Finish**. Thực ra, đến Bước 5 thì chúng ta đã có thể nhấn **Finish** để kết thúc quá trình nhập dữ liệu từ tập tin Text rồi.

Số liệu hiện ra trong SPSS sau khi được import từ tập tin **pm.txt** sẽ có dạng như sau:

HÌNH 9

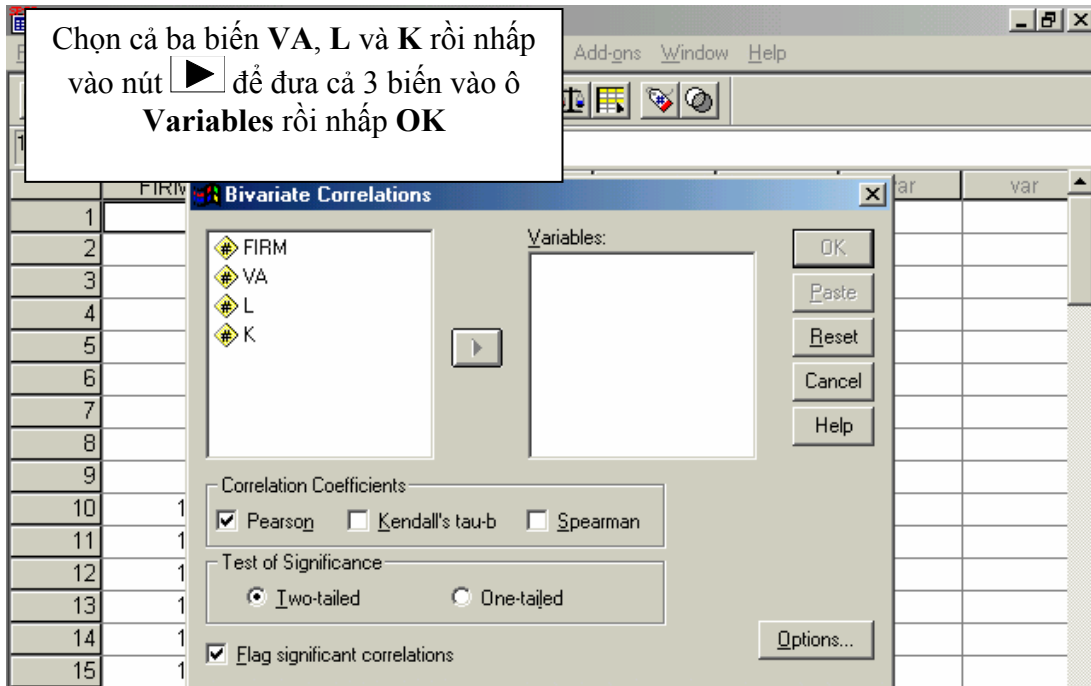


2.

Quot

Để xem xét sự tương quan giữa các biến, vào Menu **Analyze**, chọn mục **Correlation** rồi chọn **Bivariate**. Sẽ xuất hiện hình sau:

HÌNH 10



Kết quả sẽ hiện ra trong cửa sổ **Output1** như sau:

Correlations

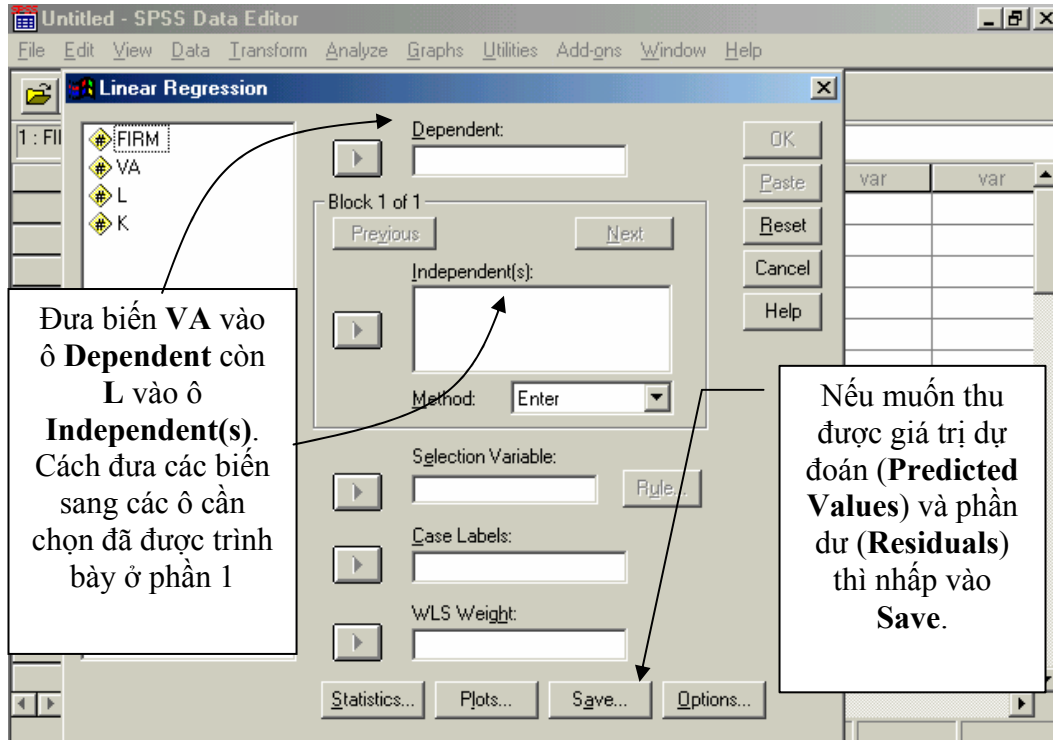
		VA	L	K
VA	Pearson Correlation	1	.965**	.975**
	Sig. (2-tailed)	.	.000	.000
	N	27	27	27
L	Pearson Correlation	.965**	1	.964**
	Sig. (2-tailed)	.000	.	.000
	N	27	27	27
K	Pearson Correlation	.975**	.964**	1
	Sig. (2-tailed)	.000	.000	.
	N	27	27	27

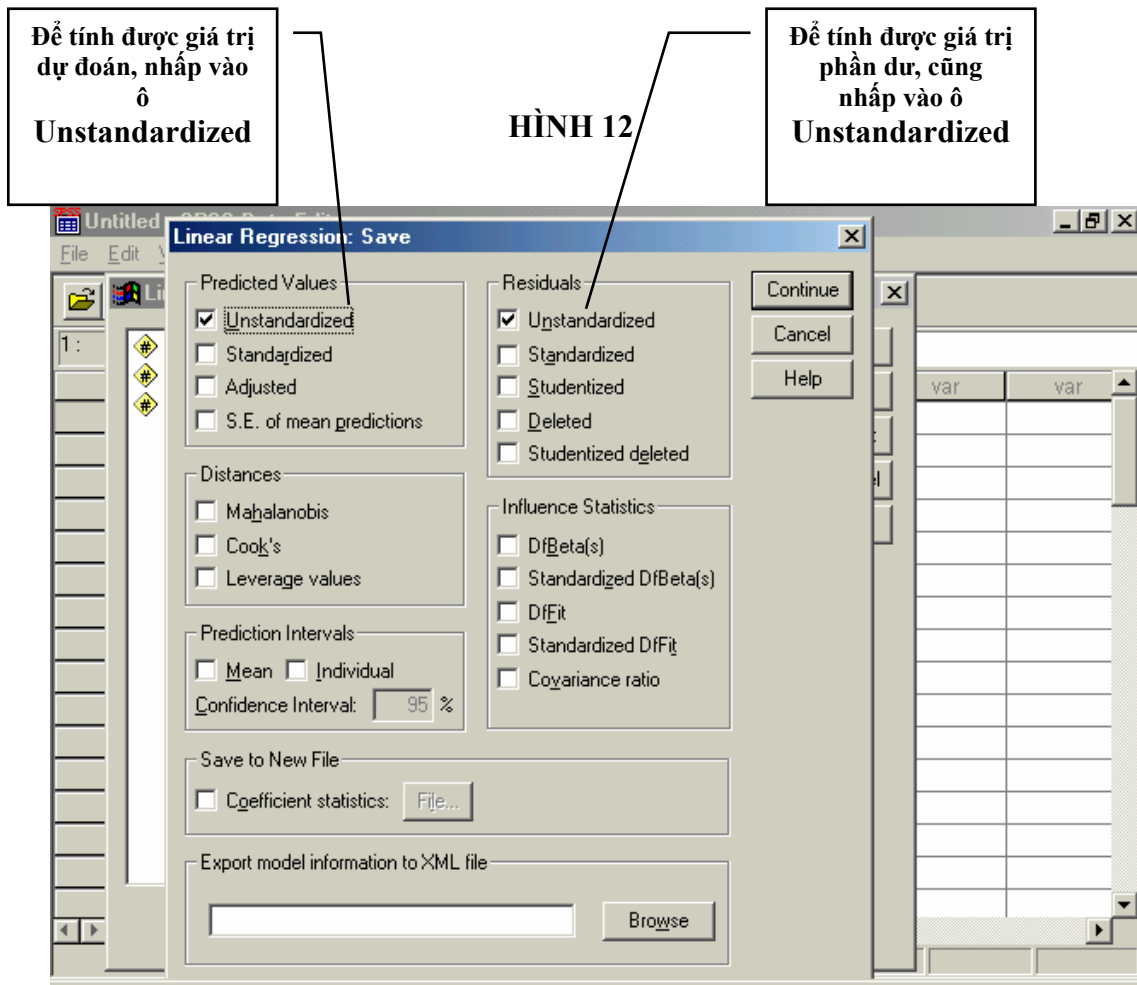
** . Correlation is significant at the 0.01 level (2-tailed).

3. Hồi quy OLS (trường hợp đơn biến)

Sau khi xem xét sự tương quan giữa các biến, chúng ta tiến hành hồi quy VA theo L. Để bắt đầu, vào menu **Analyze**, chọn mục **Regression** và chọn **Linear**.

HÌNH 11





HÌNH 12

Sau khi chọn xong thì nhấp vào Continue để trở lại hộp thoại **Linear Regression**. Để tiến hành hồi quy thì nhấp **OK**. Kết quả hồi quy sẽ được hiện ra cửa sổ **Output1** như sau:

Regression

Variables Entered/Removed^a

Model	Variables Entered	Variables Removed	Method
1	L ^a	.	Enter

a. All requested variables entered.

b. Dependent Variable: VA

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.965 ^a	.930	.928	605.74555

a. Predictors: (Constant), L

ANOVA^b

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	1.23E+08	1	122645979.7	334.251	.000 ^a
	Residual	9173192	25	366927.676		
	Total	1.32E+08	26			

a. Predictors: (Constant), L

b. Dependent Variable: VA

Coefficients^a

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	-292.397	185.269		-1.578	.127
	L	6.533	.357	.965	18.283	.000

a. Dependent Variable: VA

Trong khi đó, ở Sheet chứa dữ liệu chính, sẽ xuất hiện thêm 2 cột dữ liệu mới, cột **PRE_1** chứa giá trị dự đoán còn cột **RES_1** chứa giá trị phần dư.

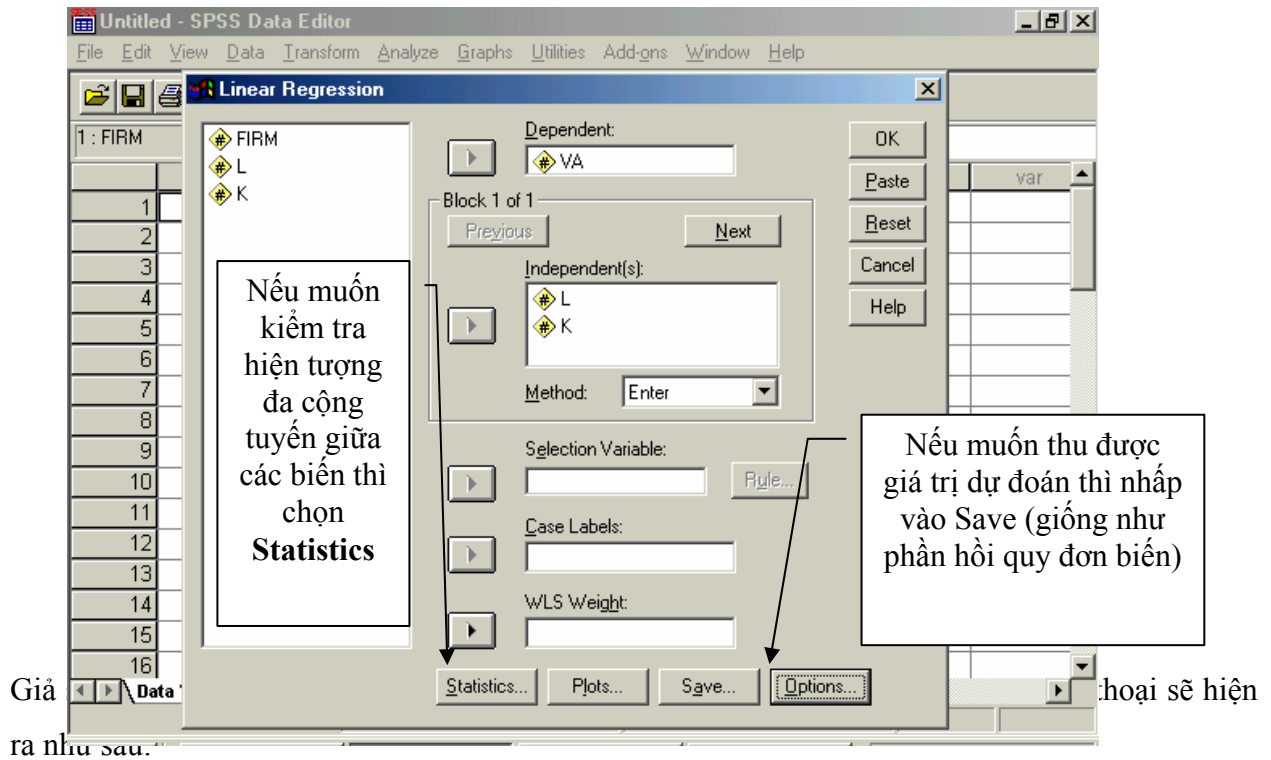
HÌNH 13

	FIRM	VA	L	K	PRE_1	RES_1	var
1	1.0	657.29	162.31	279.99	767.96723	-110.67723	
2	2.0	935.93	214.43	542.50	1108.46494	-172.53494	
3	3.0	1110.65	186.44	721.51	925.60748	185.04252	
4	4.0	1200.89	245.83	1167.68	1313.59979	-112.70979	
5	5.0	1052.68	211.40	811.77	1088.67008	-35.99008	
6	6.0	3406.02	690.61	4558.02	4219.32837	-813.30837	
7	7.0	2427.89	452.79	3069.91	2665.66056	-237.77056	
8	8.0	4257.46	714.20	5585.01	4373.44083	-115.98083	
9	9.0	1625.19	320.54	1618.75	1801.67700	-176.48700	

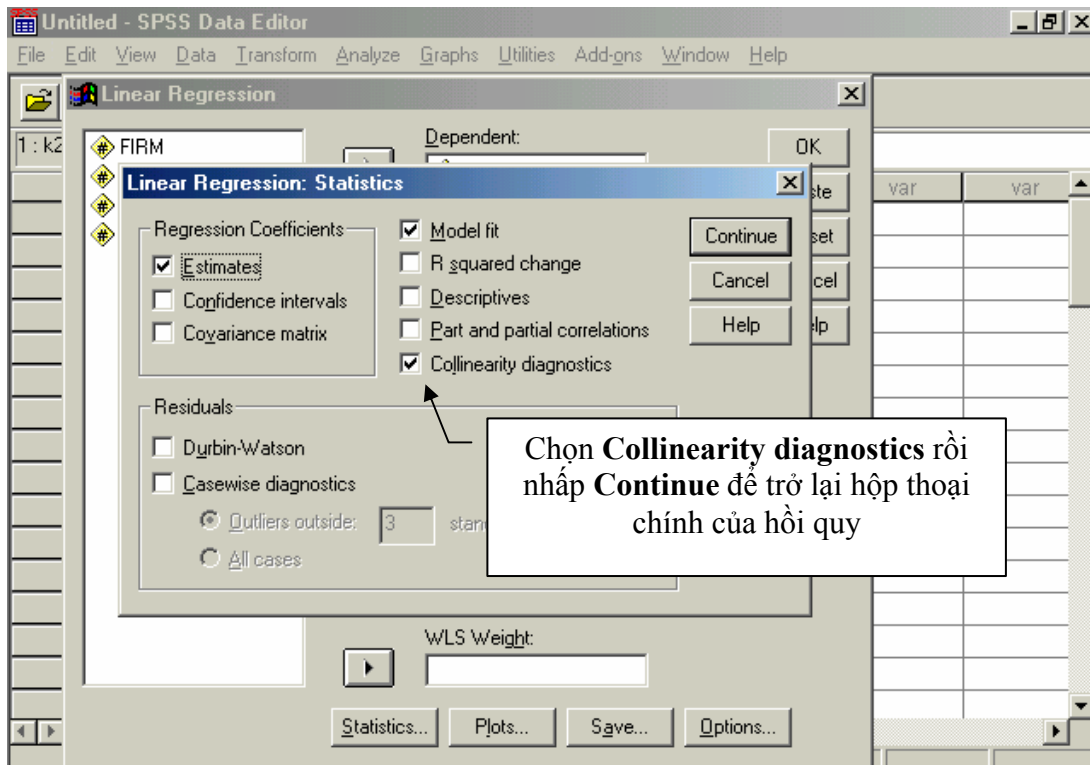
4. Hồi quy OLS (trường hợp đa biến)

Bây giờ tiến hành hồi quy VA theo K và L, cũng bắt đầu từ menu **Analyze**, chọn mục **Regression** và **Linear**. Biến VA được đưa vào ô **Dependent**, còn K và L cùng được đưa vào ô **Independent(s)**.

HÌNH 14



HÌNH 15



Sau khi trở lại hộp thoại chính của hồi quy, bấm **OK**, kết quả hồi quy đa biến như sau:

Regression

Variables Entered/Removed^b

Model	Variables Entered	Variables Removed	Method
1	K, L ^a	.	Enter

a. All requested variables entered.

b. Dependent Variable: VA

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.980 ^a	.960	.956	469.86415

a. Predictors: (Constant), K, L

ANOVA^b

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	1.27E+08	2	63260317.93	286.541	.000 ^a
	Residual	5298536	24	220772.321		
	Total	1.32E+08	26			

a. Predictors: (Constant), K, L

b. Dependent Variable: VA

Coefficients^a

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	Collinearity Statistics	
		B	Std. Error	Beta			Tolerance	VIF
1	(Constant)	114.338	173.431		.659	.516		
	L	2.338	1.039	.345	2.250	.034	.071	14.051
	K	.471	.112	.643	4.189	.000	.071	14.051

a. Dependent Variable: VA

Collinearity Diagnostics^a

Model	Dimension	Eigenvalue	Condition Index	Variance Proportions		
				(Constant)	L	K
1	1	2.619	1.000	.03	.00	.01
	2	.365	2.681	.51	.00	.03
	3	.016	12.699	.46	.99	.97

a. Dependent Variable: VA

5. Hồi quy trong trường hợp có hiện tượng đa cộng tuyến hoàn hảo

Giả sử trong dữ liệu có thêm biến $K2 = K \times 2$, thực hiện lại các bước như trong phần 4 ở trên, chúng ta thu được kết quả như sau:

Regression

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.980 ^a	.960	.956	469.86415

a. Predictors: (Constant), k2, L

ANOVA^b

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	1.27E+08	2	63260317.93	286.541	.000 ^a
	Residual	5298536	24	220772.321		
	Total	1.32E+08	26			

a. Predictors: (Constant), k2, L

b. Dependent Variable: VA

Coefficients^a

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	Collinearity Statistics	
		B	Std. Error	Beta			Tolerance	VIF
1	(Constant)	114.338	173.431		.659	.516		
	L	2.338	1.039	.345	2.250	.034	.071	14.051
	k2	.236	.056	.643	4.189	.000	.071	14.051

a. Dependent Variable: VA

Excluded Variables^b

Model	Beta In	t	Sig.	Partial Correlation	Collinearity Statistics		
					Tolerance	VIF	Minimum Tolerance
1	K	. ^a	.	.	.000	.	.000

a. Predictors in the Model: (Constant), k2, L

b. Dependent Variable: VA

Trong trường hợp này, xuất hiện 1 bảng thông báo biến bị loại bỏ ra khỏi mô hình hồi quy vì có hiện tượng đa cộng tuyến hoàn hảo, đó là biến **K**. Nguyên tắc loại biến của SPSS là loại biến bé hơn (vì $K2 = 2 \times K$).

PHÂN TÍCH DỮ LIỆU BẰNG PHẦN MỀM SPSS 12.0*

PHẦN 3

Nội dung chính trong phần này:

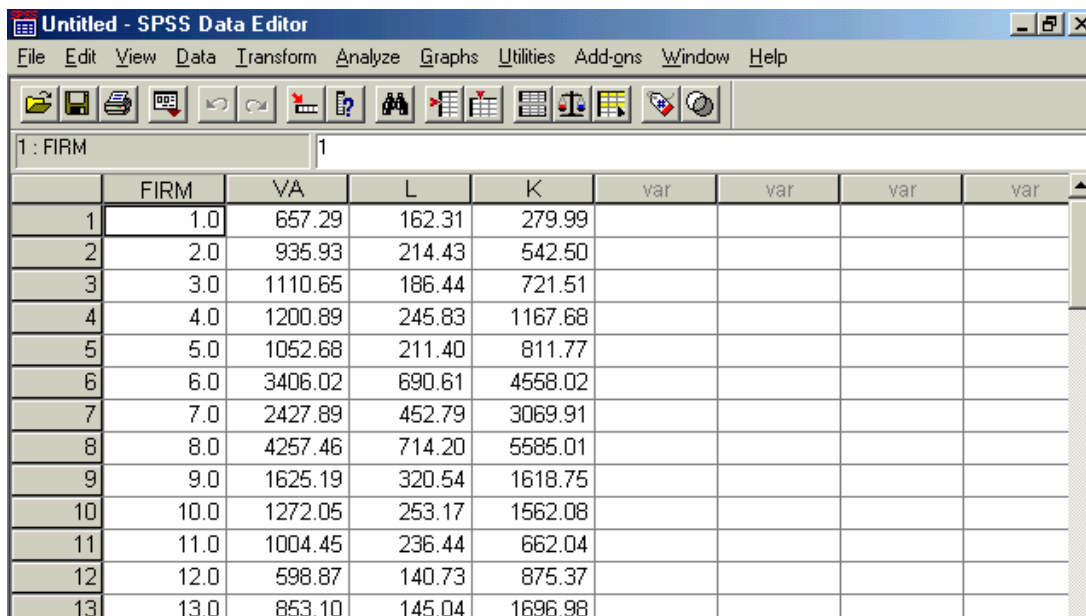
- ❖ Tạo biến mới từ các biến đã có trong dữ liệu

* SPSS 12.0 là sản phẩm đã đăng ký của SPSS Inc.

❖ **Tạo biến mới từ các biến đã có trong dữ liệu**

Bộ liệu chúng ta có như sau:

HÌNH 1

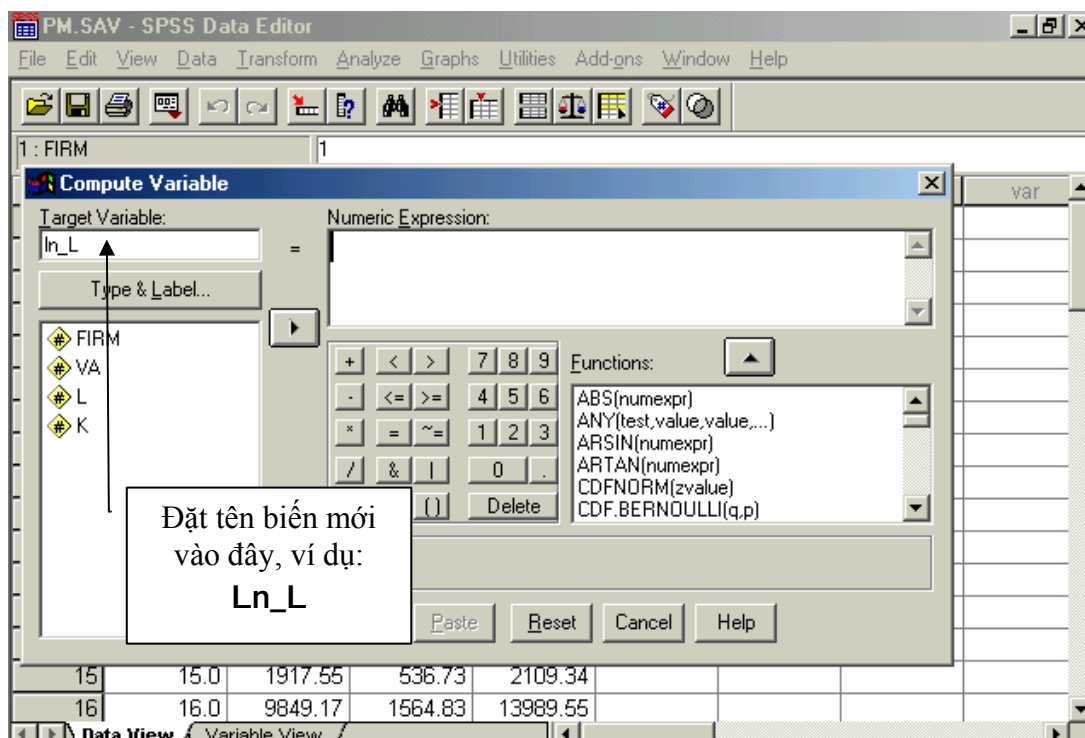


	FIRM	VA	L	K	var	var	var	var
1	1.0	657.29	162.31	279.99				
2	2.0	935.93	214.43	542.50				
3	3.0	1110.65	186.44	721.51				
4	4.0	1200.89	245.83	1167.68				
5	5.0	1052.68	211.40	811.77				
6	6.0	3406.02	690.61	4558.02				
7	7.0	2427.89	452.79	3069.91				
8	8.0	4257.46	714.20	5585.01				
9	9.0	1625.19	320.54	1618.75				
10	10.0	1272.05	253.17	1562.08				
11	11.0	1004.45	236.44	662.04				
12	12.0	598.87	140.73	875.37				
13	13.0	853.10	145.04	1696.98				

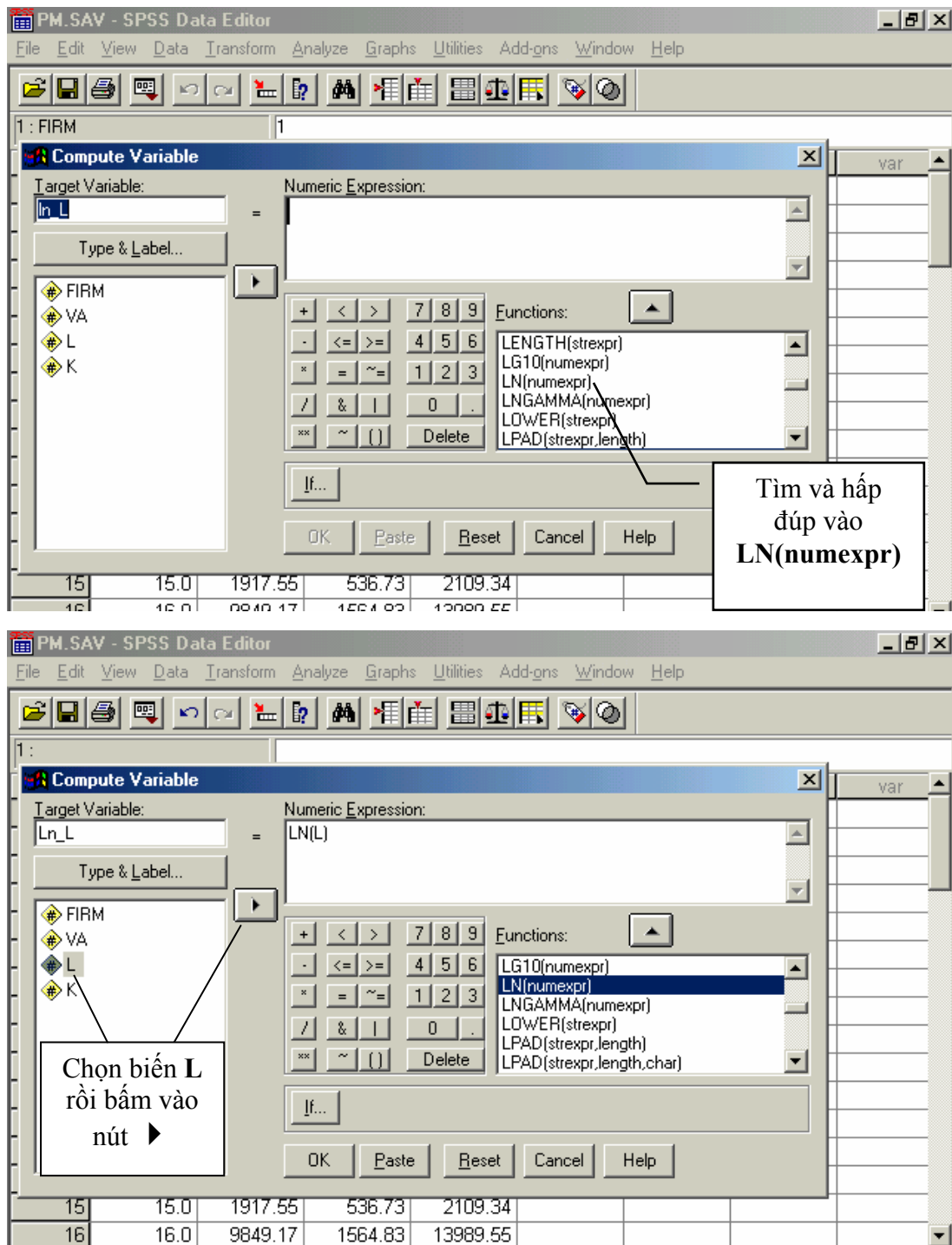
- a. Giả sử, chúng ta cần hồi quy VA theo $\ln(L)$ và $\ln(K)$, như vậy cần tạo thêm 2 biến mới là \ln_L và \ln_K từ L và K.

Để bắt đầu, chúng ta vào Menu **Transform**, chọn **Compute**.

HÌNH 2



HÌNH 3



Sau cùng, bấm **OK** để thực hiện quá trình tính toán.

HÌNH 5

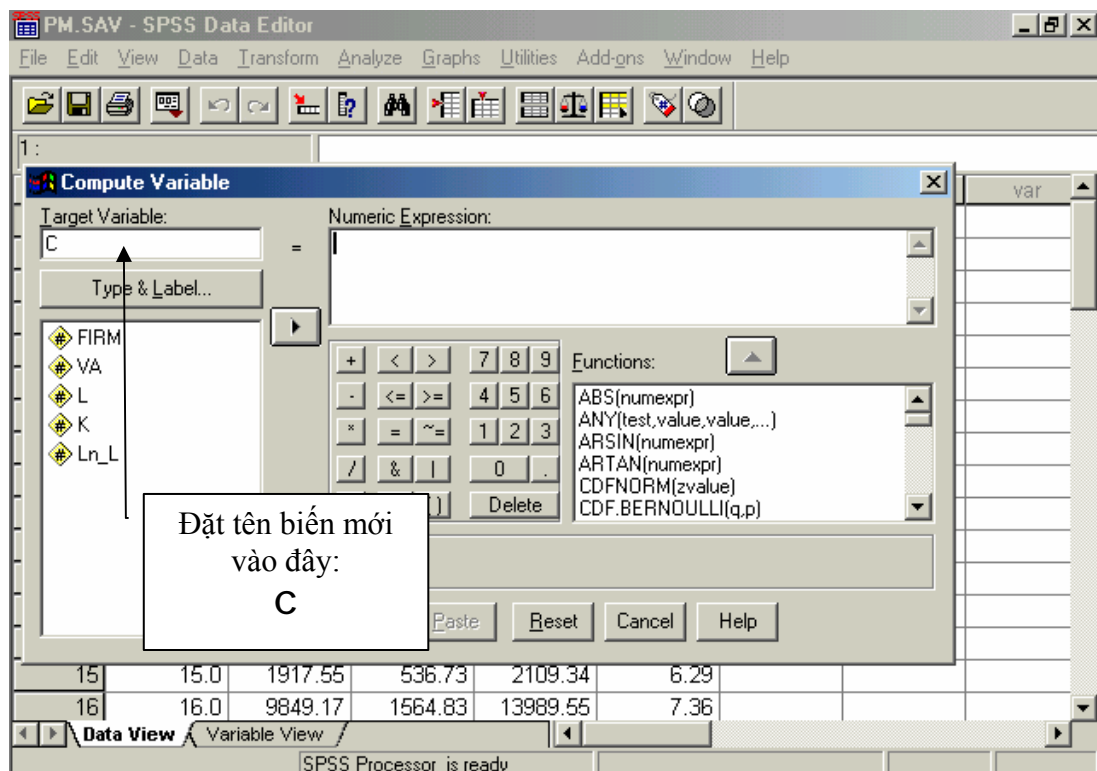
	FIRM	VA	L	K	Ln_L	var	var	var
1	1.0	657.29	162.31	279.99	5.09			
2	2.0	935.93	214.43	542.50	5.37			
3	3.0	1110.65	186.44	721.51	5.23			
4	4.0	1200.89	245.83	1167.68	5.50			
5	5.0	1052.68	211.40	811.77	5.35			
6	6.0	3406.02	690.61	4558.02	6.54			
7	7.0	2427.89	452.79	3069.91	6.12			
8	8.0	4257.46	714.20	5585.01	6.57			
9	9.0	1625.19	320.54	1618.75	5.77			
10	10.0	1272.05	253.17	1562.08	5.53			
11	11.0	1004.45	236.44	662.04	5.47			
12	12.0	598.87	140.73	875.37	4.95			

Để tạo ra biến **Ln_K** thì các bước thực hiện cũng tương tự.

- b. Giả sử, chúng ta cần hồi quy **VA** theo **L** và **C** với $C = 2K+1$. Như vậy cần tạo thêm 2 biến mới là **C**.

Để bắt đầu, chúng ta cũng vào Menu **Transform**, chọn **Compute**.

HÌNH 6



HÌNH 7

The screenshot shows the SPSS Data Editor interface. The 'Compute Variable' dialog box is open, with 'C' as the target variable and 'K' as the numeric expression. The variable list on the left includes FIRM, VA, L, K, and Ln L. A callout box points to the 'K' variable, indicating it should be selected. Another callout box points to the 'Numeric Expression' field, indicating that the user should enter '*2+1' and click 'OK'. Below the dialog box, a data table is visible with columns FIRM, VA, L, K, Ln L, and C. A callout box points to the 'C' column, indicating that the variable has been added to the data.

	FIRM	VA	L	K	Ln L	C
1	1.0	657.29	162.31	279.99	5.09	560.98
2	2.0	935.93	214.43	542.50	5.37	1086.00
3	3.0	1110.65	186.44	721.51	5.23	1444.02
4	4.0	1200.89	245.83	1167.68	5.50	2336.36
5	5.0	1052.68	211.40	811.77	5.35	1624.54
6	6.0	3406.02	690.61	4558.02	6.54	9117.04
7	7.0	2427.89	452.79	3069.91	6.12	6140.82
8	8.0	4257.46	714.20	5585.01	6.57	11171.02
9	9.0	1625.19	320.54	1618.75	5.77	3238.50
10	10.0	1272.05	253.17	1562.08	5.53	3125.16
11	11.0	1004.45	236.44	662.04	5.47	1325.08
12	12.0	598.87	140.73	875.37	4.95	1751.74
13	13.0	853.10	145.04	1696.98	4.98	3394.96
14	14.0	1165.63	240.27	1078.79	5.48	2158.58
15	15.0	1917.55	536.73	2109.34	6.29	4219.68
16	16.0	9849.17	1564.83	13989.55	7.36	27980.10

PHÂN TÍCH DỮ LIỆU BẰNG PHẦN MỀM SPSS 12.0*

PHẦN 4

Nội dung chính trong phần này:

1. Khai báo các thông số của biến
2. Tạo biến giả
3. Hồi quy OLS kết hợp với phương pháp **Stepwise**

* SPSS 12.0 là sản phẩm đã đăng ký của SPSS Inc.

1. Khai báo các thông số của các biến trong bộ dữ liệu

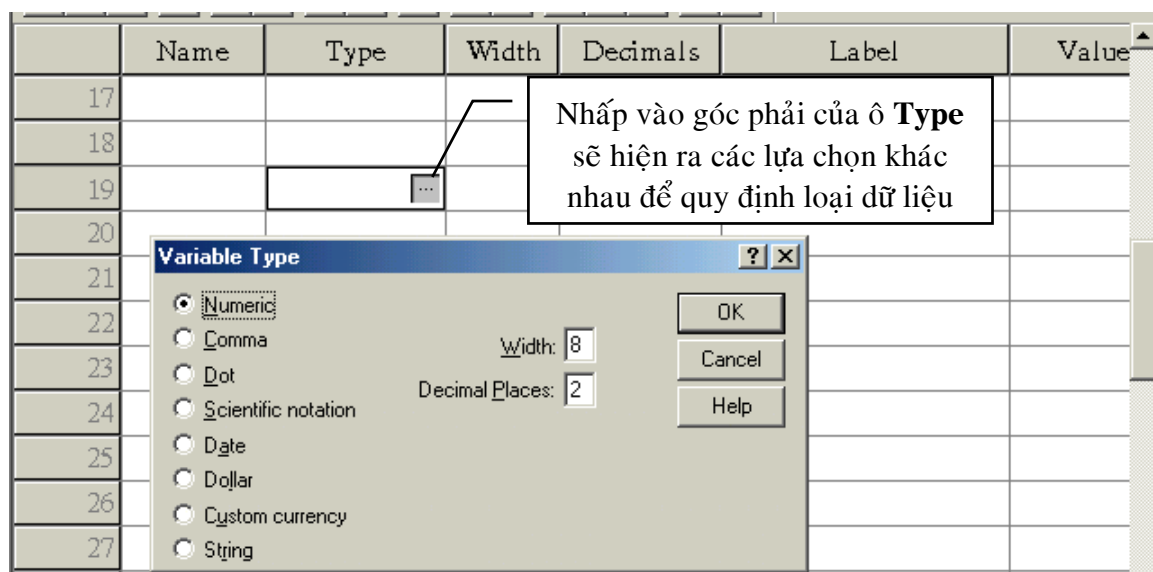
Khi sử dụng SPSS 12.0, chúng ta thường thấy hai Sheet: **Data View** và **Variable View**. **Data View** chứa dữ liệu còn **Variable View** chứa các thông tin của các biến trong dữ liệu. Các thông tin này bao gồm:

- Name:** tên của biến.
- Type:** loại dữ liệu của biến.
- Width:** số lượng ký tự hay số lượng chữ số được hiển thị.
- Decimals:** số lượng chữ số thập phân.
- Label:** nhãn của biến.
- Values:** nhãn hoặc giá trị của các quan sát trong biến (phát huy tác dụng tốt trong thống kê mô tả).
- Missing:** số lượng quan sát bị khuyết.
- Columns:** chiều rộng của cột.
- Align:** vị trí (nếu là số thì sẽ là bên phải, còn là ký tự sẽ nằm bên trái)
- Measure:** thang đo

Chi tiết cho một số thông tin quan trọng của biến:

- Name:** tên của biến. Tên biến phải bắt đầu bằng một chữ và có độ dài tối đa là 64 ký tự (không sử dụng các ký tự đặc biệt, không kết thúc tên biến bằng dấu chấm “.”)
- Type:** loại dữ liệu của biến.

HÌNH 1



c. **Measure:** thang đo

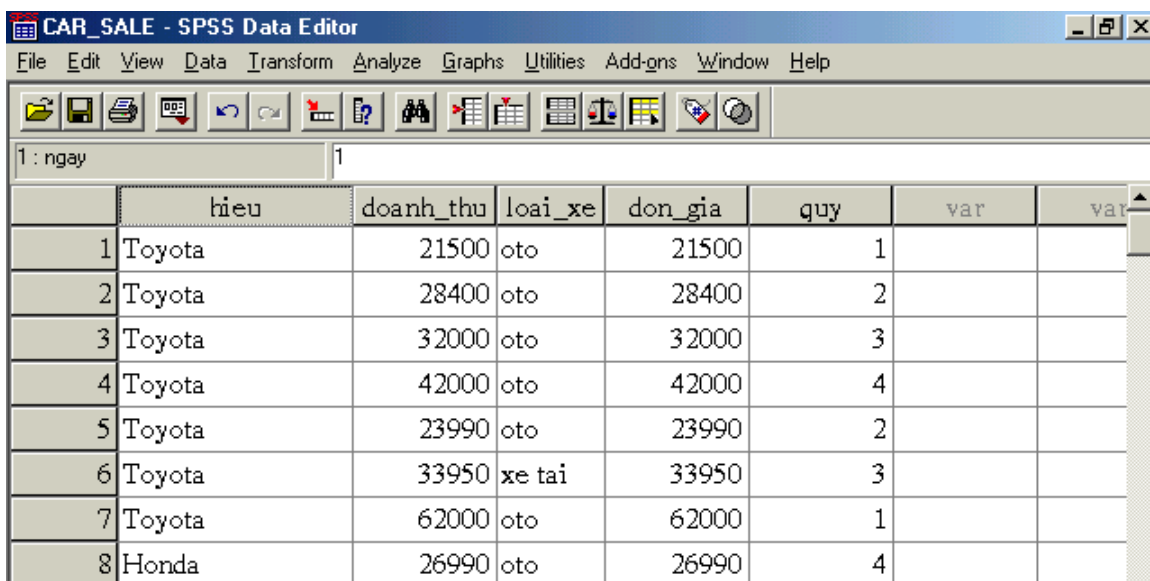
Trong SPSS 12.0 có 3 loại thang đo: Scale, Nominal và Ordinal.

- **Scale:** cho biết dữ liệu là những con số định lượng (ví dụ: thu nhập, tuổi, chiều cao ...).
- **Nominal:** dữ liệu là chữ hoặc con số định tính (ví dụ: nam, nữ, hay 0, 1).
- **Ordinal:** dữ liệu là chữ hoặc con số định tính **nhưng chú ý đến thứ bậc, mức độ cao thấp, nặng nhẹ...** (ví dụ: thấp, trung bình, cao; kịch liệt phản đối, phản đối, nhất trí, nhất trí cao).

2. **Tạo biến giả**

Giả sử chúng ta có bộ dữ liệu sau được import từ Excel:

HÌNH 2



	hieu	doanh_thu	loai_xe	don_gia	quy	var	var
1	Toyota	21500	oto	21500	1		
2	Toyota	28400	oto	28400	2		
3	Toyota	32000	oto	32000	3		
4	Toyota	42000	oto	42000	4		
5	Toyota	23990	oto	23990	2		
6	Toyota	33950	xe tai	33950	3		
7	Toyota	62000	oto	62000	1		
8	Honda	26990	oto	26990	4		

Dữ liệu này là các quan sát ngẫu nhiên của một cửa hàng bán ô tô và xe tải trong năm.

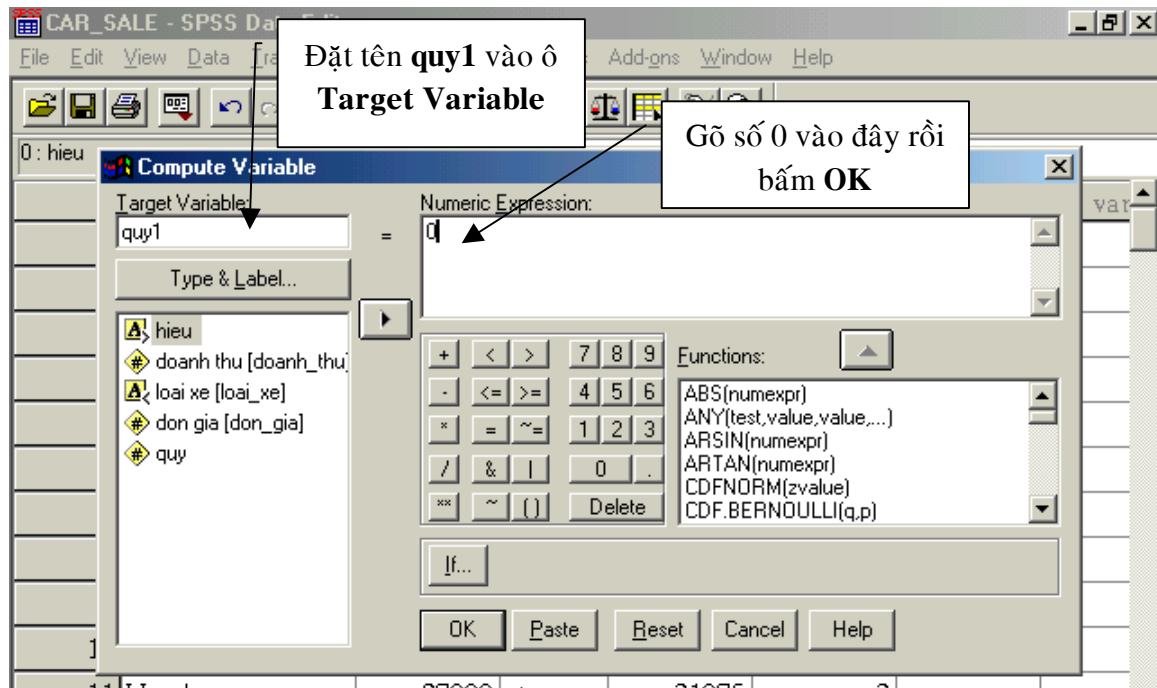
Trong đó:

- hieu: tên của nhà sản xuất.
- doanh_thu: doanh thu trong ngày quan sát (USD).
- loai_xe: loại xe ô tô hay xe tải.
- don_gia: đơn giá (USD).
- quy: quý mà quan sát đổ rơi vào.

2.1. Tạo 3 biến giả thể hiện Quý 1, Quý 2 và Quý 3

Vào Menu **Transform**, chọn **Compute**

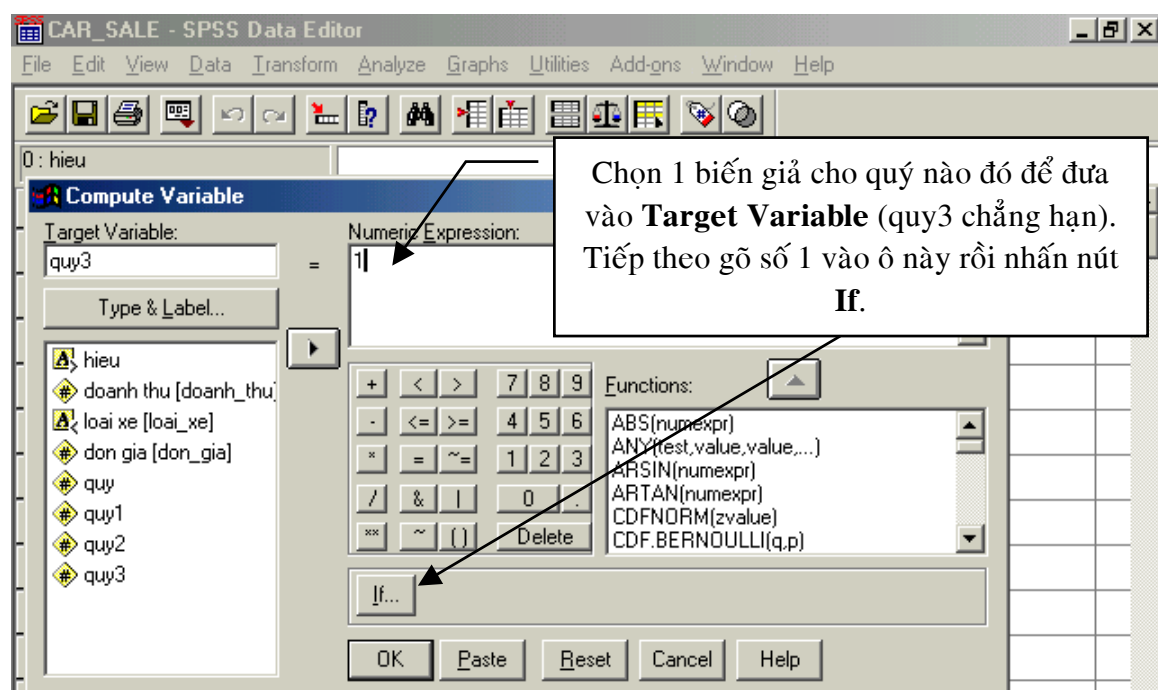
HÌNH 3



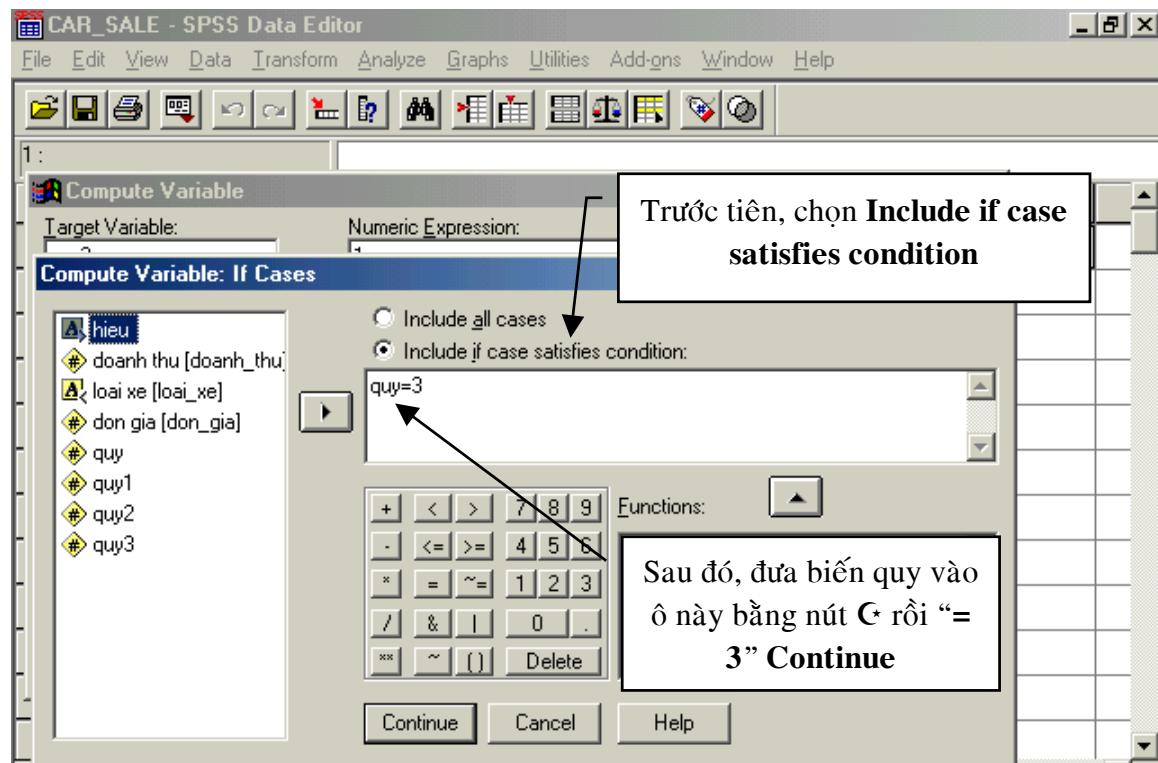
Lần lượt thực hiện như thế cho **quy2** và **quy3**. Lúc này trong dữ liệu đã xuất hiện biến **quy1**, **quy2** và **quy3** với tất cả các giá trị đều bằng 0.

Tiếp theo, lại trở vào **Compute**.

HÌNH 4



HÌNH 5

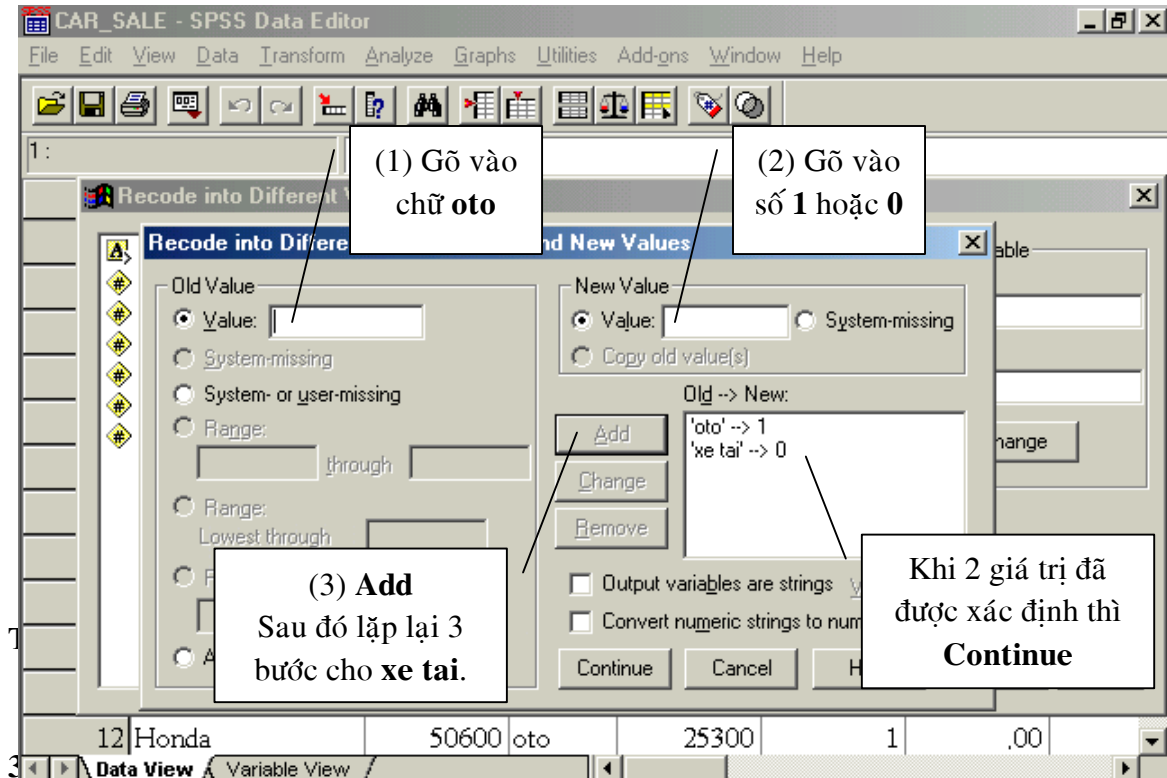
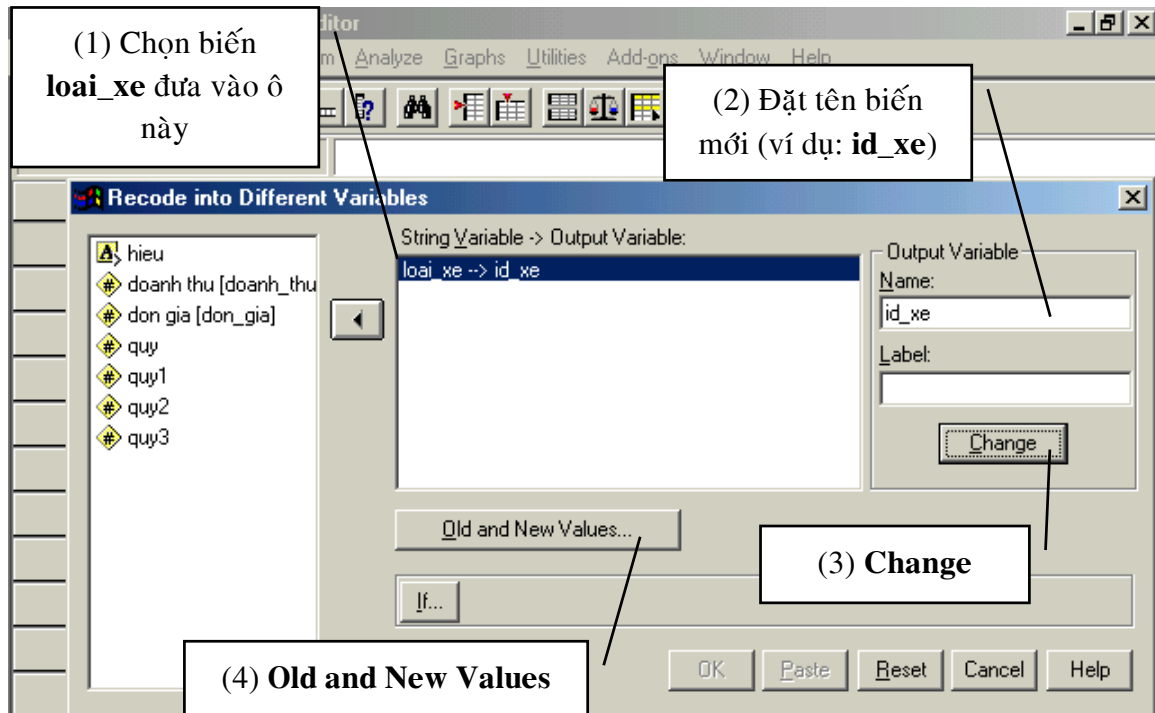


Trở lại hộp thoại phía trước, tiếp tục chọn OK để hoàn tất tạo biến giả **quy3**, và tiếp tục làm tương tự cho quy1 và quy2.

2.2. Tạo biến giả cho loại xe

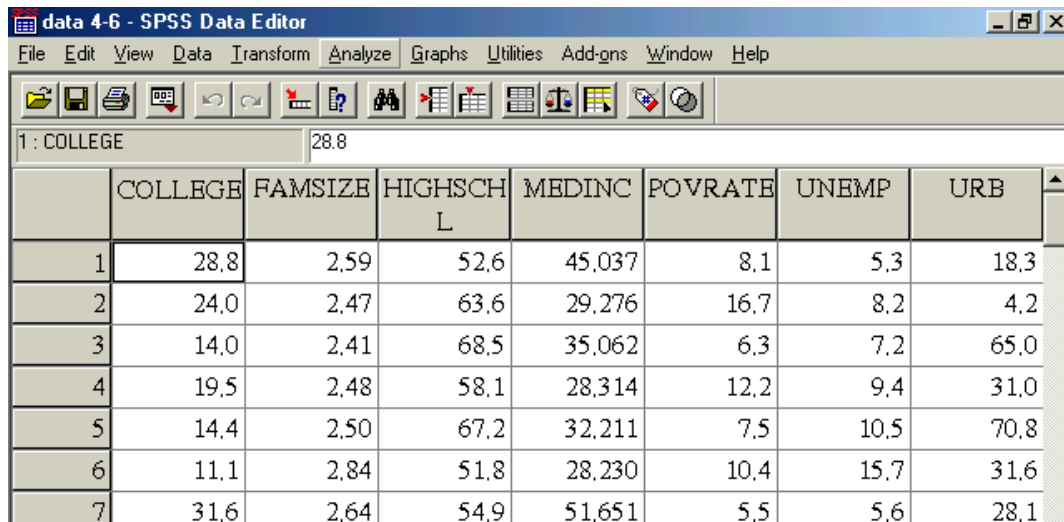
Vào **Transform, Recode, Into Different Variables**. Tức là chúng ta sẽ mã hóa lại biến **loai_xe**, và sẽ cho ra một biến mới (nếu chọn **Into Same Variables** thì SPSS sẽ biến đổi rồi thay thế luôn thông tin của biến cũ).

HÌNH 6



Lấy dữ liệu từ file DATA4-6 của Ramanathan.

HÌNH 8



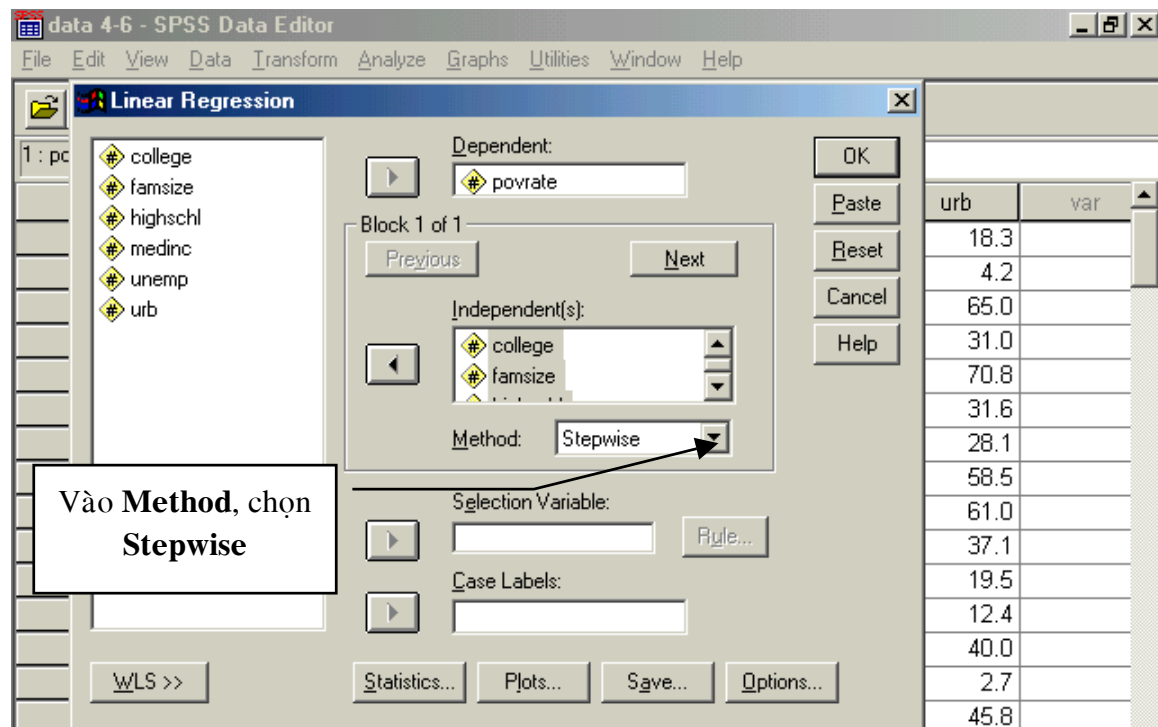
	COLLEGE	FAMSIZE	HIGHSCH L	MEDINC	POVRATE	UNEMP	URB
1	28,8	2,59	52,6	45,037	8,1	5,3	18,3
2	24,0	2,47	63,6	29,276	16,7	8,2	4,2
3	14,0	2,41	68,5	35,062	6,3	7,2	65,0
4	19,5	2,48	58,1	28,314	12,2	9,4	31,0
5	14,4	2,50	67,2	32,211	7,5	10,5	70,8
6	11,1	2,84	51,8	28,230	10,4	15,7	31,6
7	31,6	2,64	54,9	51,651	5,5	5,6	28,1

Bây giờ chúng ta sẽ hồi quy OLS kết hợp với phương pháp **Stepwise** với biến phụ thuộc là POV RATE, biến độc lập là tất cả các biến còn lại trong dữ liệu.

Tác dụng của phương pháp **Stepwise** được hiểu nôm na là giúp chúng ta tìm ra được những kết hợp của các biến độc lập sao cho kết quả hồi quy sẽ “tốt” theo hướng các giá trị thống kê t , F có ý nghĩa, và việc lựa chọn các kết hợp này sẽ được căn cứ vào khả năng làm gia tăng giá trị của R^2 .

Để bắt đầu, vào Menu **Analyze, Regression, Linear** rồi đưa biến POV RATE và ô **Dependent** và các biến còn lại vào **Independent(s)**.

HÌNH 9



Kết quả hồi quy được trình bày như sau:

Bảng 1: Trình bày thông tin cho biết SPSS đã tìm ra được bao nhiêu kết hợp tốt theo thống kê t và F. Đồng thời, các mô hình xuất hiện sau sẽ có giá trị R^2 và R^2 hiệu chỉnh lớn hơn mô hình xuất hiện trước (xem bảng 2).

Variables Entered/Removed(a)

Model	Variables Entered	Variables Removed	Method
1	MEDINC	.	Stepwise (Criteria: Probability-of-F-to-enter <= .050, Probability-of-F-to-remove >= .100).
2	HIGHSCHL	.	Stepwise (Criteria: Probability-of-F-to-enter <= .050, Probability-of-F-to-remove >= .100).
3	FAMSIZE	.	Stepwise (Criteria: Probability-of-F-to-enter <= .050, Probability-of-F-to-remove >= .100).
4	COLLEGE	.	Stepwise (Criteria: Probability-of-F-to-enter <= .050, Probability-of-F-to-remove >= .100).

a. Dependent Variable: POV RATE

Bảng 2:

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.782(a)	.612	.605	2.4870
2	.895(b)	.800	.793	1.7999
3	.903(c)	.816	.805	1.7445
4	.912(d)	.831	.818	1.6870

a Predictors: (Constant), MEDINC

b Predictors: (Constant), MEDINC, HIGHSCHL

c Predictors: (Constant), MEDINC, HIGHSCHL, FAMSIZE

d Predictors: (Constant), MEDINC, HIGHSCHL, FAMSIZE, COLLEGE

Bảng 3:

ANOVA(e)

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	545.424	1	545.424	88.181	.000(a)
	Residual	346.376	56	6.185		
	Total	891.799	57			
2	Regression	713.626	2	356.813	110.144	.000(b)
	Residual	178.173	55	3.240		
	Total	891.799	57			
3	Regression	727.461	3	242.487	79.679	.000(c)
	Residual	164.338	54	3.043		
	Total	891.799	57			
4	Regression	740.961	4	185.240	65.088	.000(d)
	Residual	150.838	53	2.846		
	Total	891.799	57			

a Predictors: (Constant), MEDINC

b Predictors: (Constant), MEDINC, HIGHSCHL

c Predictors: (Constant), MEDINC, HIGHSCHL, FAMSIZE

d Predictors: (Constant), MEDINC, HIGHSCHL, FAMSIZE, COLLEGE

e Dependent Variable: POVRATE

Bảng 4: Các hệ số hồi quy và thống kê t

Coefficients(a)

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	23.131	1.446		15.997	.000
	MEDINC	-.374	.040	-.782	-9.390	.000
2	(Constant)	41.849	2.801		14.943	.000
	MEDINC	-.435	.030	-.909	-14.475	.000
3	HIGHSCHL	-.288	.040	-.452	-7.206	.000
	(Constant)	31.775	5.449		5.831	.000
	MEDINC	-.421	.030	-.880	-14.131	.000
4	HIGHSCHL	-.235	.046	-.369	-5.111	.000
	FAMSIZE	2.434	1.141	.148	2.132	.038
	(Constant)	19.172	7.826		2.450	.018
	MEDINC	-.552	.067	-1.154	-8.284	.000
	HIGHSCHL	-.139	.063	-.218	-2.214	.031
	FAMSIZE	5.414	1.758	.329	3.079	.003
	COLLEGE	.195	.090	.380	2.178	.034

a Dependent Variable: POVRATE

Bảng 5: Các biến bị bỏ ra trong quá trình chạy hồi quy

Excluded Variables(e)

Model		Beta In	T	Sig.	Partial Correlation	Collinearity Statistics
						Tolerance
1	COLLEGE	.157(a)	.998	.323	.133	.281
	FAMSIZE	.339(a)	4.809	.000	.544	.999
	HIGHSCHL	-.452(a)	-7.206	.000	-.697	.921
	UNEMP	.342(a)	3.082	.003	.384	.490
	URB	-.094(a)	-1.133	.262	-.151	.993
2	COLLEGE	-.038(b)	-.324	.747	-.044	.266
	FAMSIZE	.148(b)	2.132	.038	.279	.708
	UNEMP	.071(b)	.733	.467	.099	.386
	URB	-.010(b)	-.155	.878	-.021	.955
3	COLLEGE	.380(c)	2.178	.034	.287	.105
	UNEMP	-.055(c)	-.485	.630	-.066	.270
	URB	-.114(c)	-1.620	.111	-.217	.666
4	UNEMP	.025(d)	.212	.833	.029	.242
	URB	-.091(d)	-1.296	.201	-.177	.645

a Predictors in the Model: (Constant), MEDINC

b Predictors in the Model: (Constant), MEDINC, HIGHSCHL

c Predictors in the Model: (Constant), MEDINC, HIGHSCHL, FAMSIZE

d Predictors in the Model: (Constant), MEDINC, HIGHSCHL, FAMSIZE, COLLEGE

e Dependent Variable: POVRATE

Tuy nhiên, việc lựa chọn mô hình thích hợp cho nghiên cứu còn phụ thuộc vào nhiều yếu tố khác nữa, phương pháp Stepwise chỉ là một cách giúp có thể chúng ta tiết kiệm thời gian hay gợi ra một ý tưởng về việc kết hợp các biến độc lập trong quá trình hồi quy. Nếu như chúng ta chưa nắm vững về hồi quy bội thì sẽ không phát huy được tiện ích của phương pháp này.

TRƯỜNG ĐẠI HỌC NHA TRANG
KHOA KINH TẾ
BỘ MÔN KINH TẾ CƠ SỞ



PHÂN TÍCH DỮ LIỆU NGHIÊN CỨU VỚI PHẦN
MỀM EViews 5.1

Nha Trang, tháng 03 năm 2009

TRƯỜNG ĐẠI HỌC NHA TRANG
KHOA KINH TẾ
BỘ MÔN KINH TẾ CƠ SỞ

Th.S. PHẠM THÀNH THÁI

**PHÂN TÍCH DỮ LIỆU NGHIÊN CỨU VỚI PHẦN
MỀM EVIEWS 5.1**

Nha Trang, tháng 03 năm 2009

Giới thiệu

Có rất nhiều phần mềm xử lý dữ liệu cho môn học Kinh tế lượng. Các phần mềm thường được sử dụng có các tên gọi như là STATA, SPSS, EXCEL, MINITAB, và Eviews. . . Các phần mềm này đều có điểm chung là giúp chúng ta xử lý dữ liệu một cách nhanh chóng. Tuy nhiên, mỗi phần mềm lại có những đặc điểm riêng. STATA có thể tốt cho các dữ liệu từ các cuộc điều tra lớn, SPSS có ưu điểm xử lý dữ liệu mô tả tốt dưới dạng bảng biểu, EXCEL thì có ở khắp mọi máy tính PC thông thường mà không cần phải cài đặt gì thêm... Riêng môn học kinh tế lượng, chúng ta sẽ chủ yếu sử dụng phần mềm Eviews với phiên bản 5.1. Ưu điểm chính của EViews là có thể cho chúng ta kết quả nhanh chóng về hàm kinh tế lượng cho các dữ liệu chéo (Cross-Section data), dữ liệu chuỗi thời gian (Time series data) và dữ liệu bảng (Panel data). Ngoài ra, phần mềm này lại được chạy trong môi trường Window nên rất ít khi cần nhớ các lệnh cụ thể.

Chương 1: LÀM VIỆC VỚI EVIEWS 5.1

1.1. Khởi động Eviews.

Biểu tượng của EViews trên màn hình Windows trông như thế này:



Nhấn đúp vào biểu tượng **EViews** và EViews bắt đầu hoạt động. Hãy ghi nhận menu chính, cửa sổ lệnh, cửa sổ chính, và dòng tình trạng (status line).

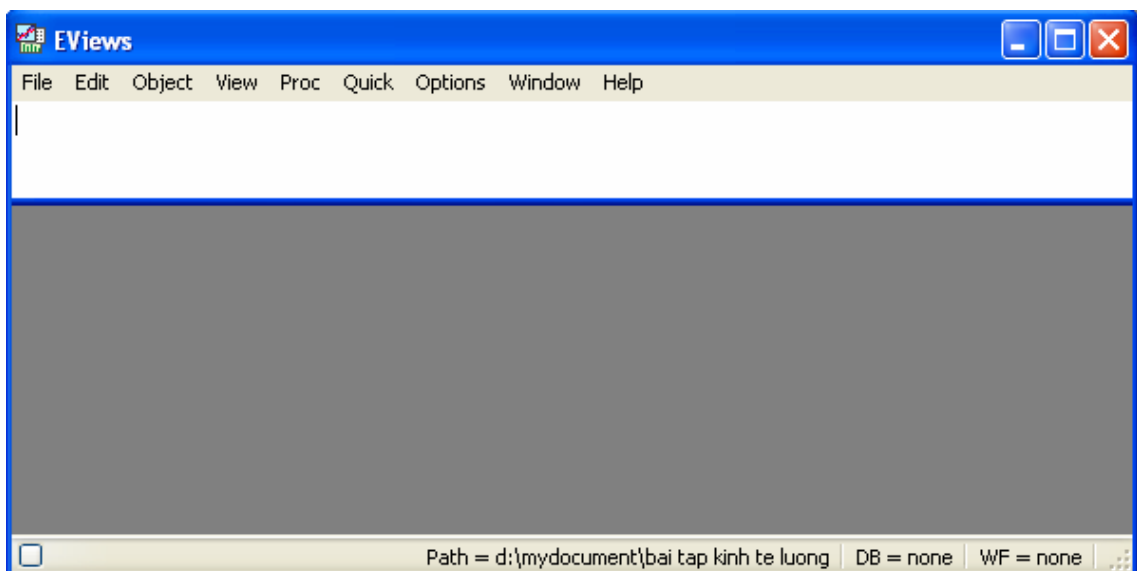
Menu chính bao gồm những lựa chọn sau:

File Edit Objects View Procs Quick Options Window Help

Hãy nhấp vào mỗi một trong những lựa chọn này và kiểm tra các menu phụ xuất hiện ở phía dưới. Các bạn có thể muốn tìm hiểu một số chi tiết về các chủ đề trong **Help**. Phương tiện **Help** của **Eviews** rất tuyệt vời.

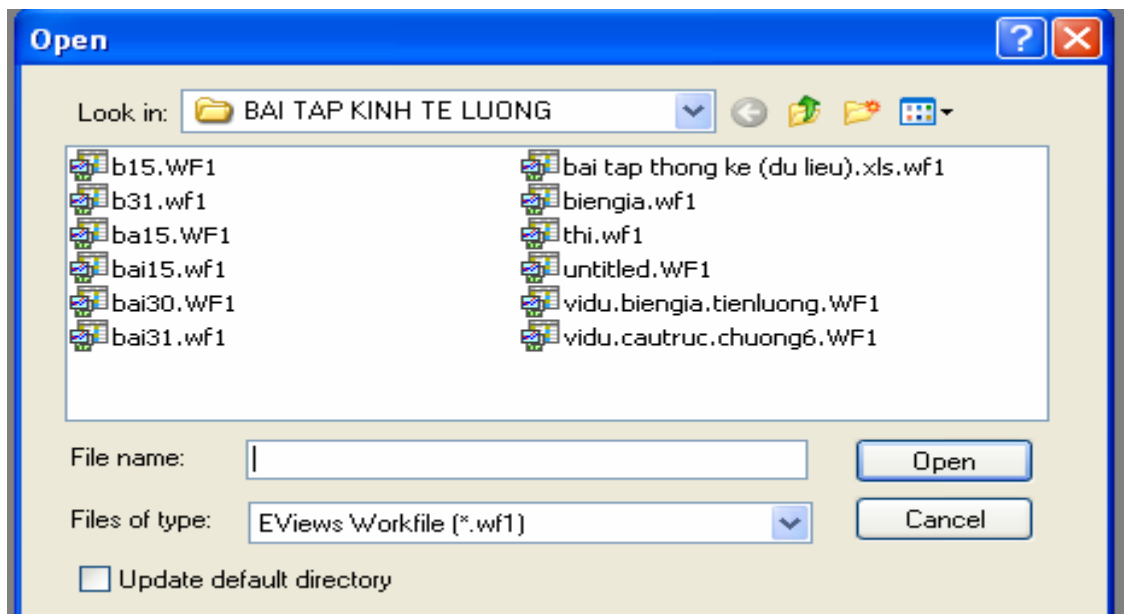
Dưới menu chính này là một cửa sổ, người ta gọi là cửa sổ lệnh. Cửa sổ này dùng để viết các câu lệnh và thực hiện một số công việc khác.

Khi các bạn mở **EViews** lần đầu tiên, cửa sổ chính còn trống vì các bạn còn chưa xác định tập tin làm việc nào (workfile) để sử dụng.



1.2. Mở tập tin làm việc (Workfile) đã lưu giữ từ trước.

Nhấp **File/Open/Workfile**; một màn hình mở ra, trên đó liệt kê các tập tin trong thư mục mặc định (default folder). Nếu thư mục mặc định không phải là thư mục mà anh chị mong muốn, thì các bạn có thể tìm trong các thư mục cho tới khi tìm ra thư mục mà các bạn muốn dùng.

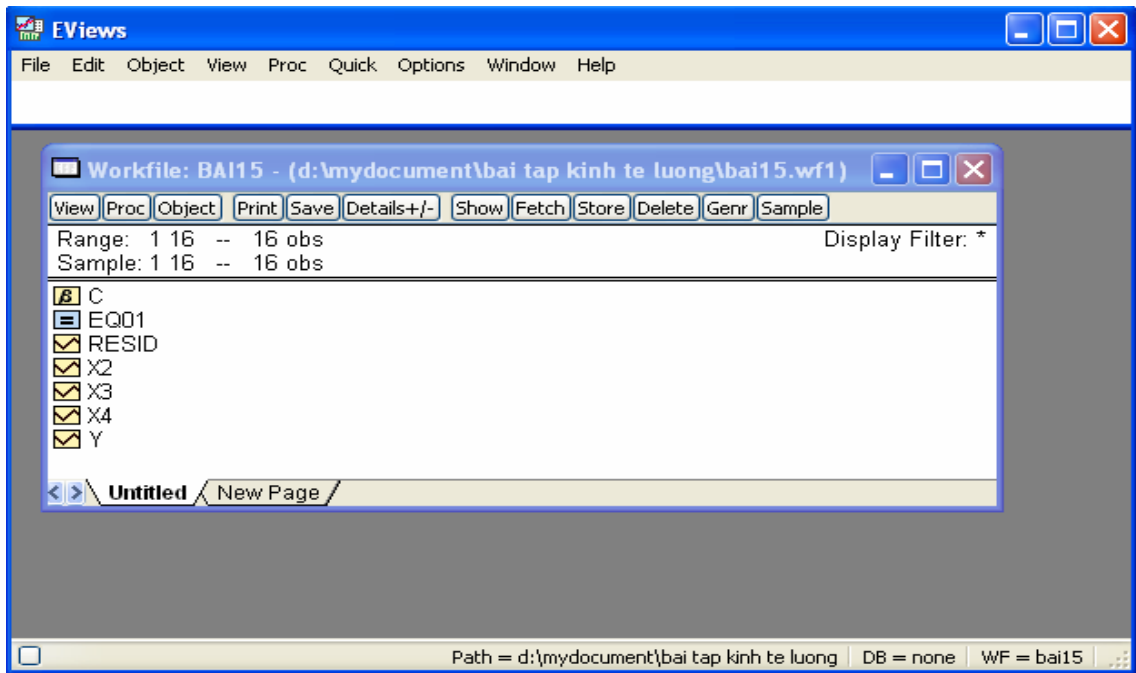


Để chỉ ra các workfile trên **EViews**, dòng "Files of type" cần xác định **Workfile (*.wf1)**.

Vì các bạn dự định sử dụng thư mục này thường xuyên, nên các bạn nên nhấp chuột để đánh dấu vào ô vuông (**Update default directory**) giúp cập nhật thư mục mặc định ở góc trái-phía dưới cửa sổ Open như trên đây. Lần sau, khi các bạn khởi động **EViews**, chuỗi **File/Open/Workfile** sẽ tự động chỉ tới thư mục này.

Mở workfile có tên **bai15.wf1** bằng cách nhấp đúp vào nó. Cách khác, các bạn có thể bôi đen nó bằng cách nhấp đơn, sau đó nhấp Open.

Với workfile đang mở, màn hình của các bạn có dạng:



Thanh trên cùng trên menu của Workfile chỉ toàn bộ đường dẫn đối với Workfile này; thanh tình trạng ở dưới đáy của màn hình cũng chỉ đường dẫn này, và nó chỉ workfile đang sử dụng : WF = bai15.

Menu của workfile chứa các nút bấm, bao gồm:

View Procs Objects Label +/- Show Fetch Store Delete Genr Sample

Các phím bấm với +/- là các phím chuyển đổi thực hiện các chức năng bật và tắt.

Thông tin về workfile này xuất hiện dưới thanh menu:

Range

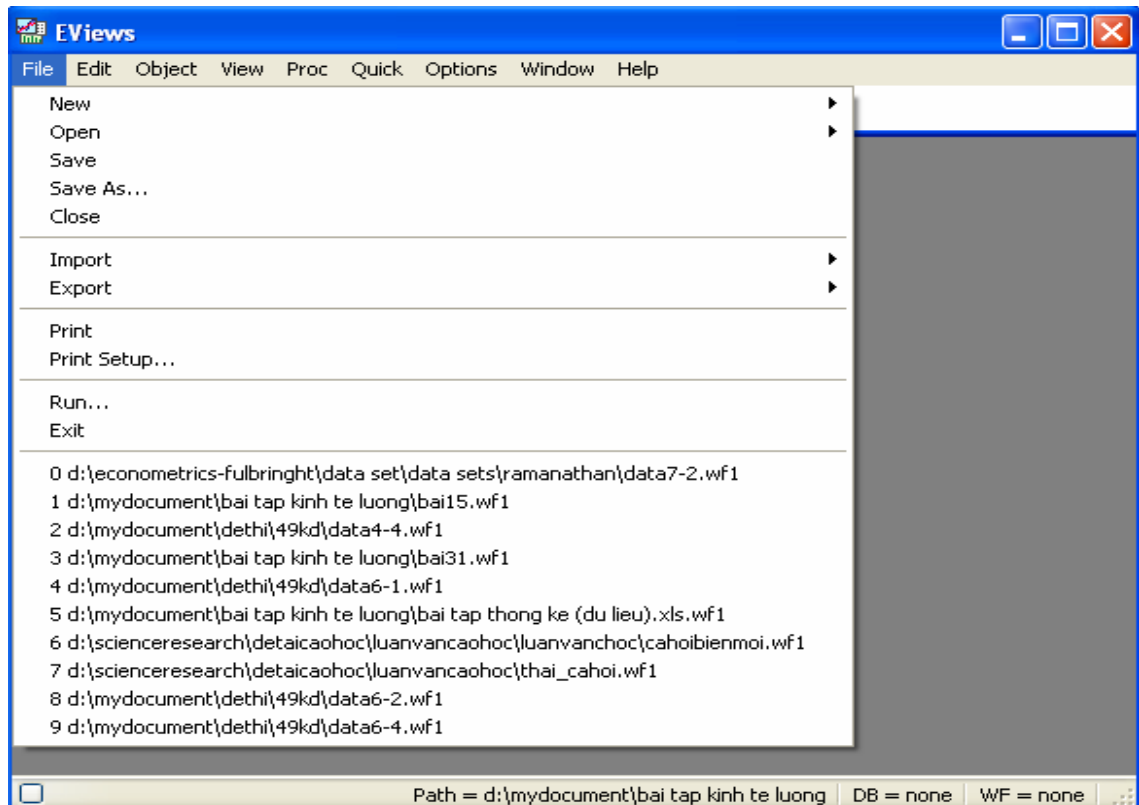
Display Filter: *

Sample

Cuối cùng, phía trong cửa sổ của workfile, chúng ta thấy tất cả các đối tượng đang có mặt trong workfile này: các chuỗi (các biến), các nhóm chuỗi, vector hệ số, vector phần dư, và bất cứ phương trình, đồ thị hay bảng nào đã được đặt tên.

Muốn đóng file **bai15.wf1** hãy nhấp vào dấu **X** ở góc phải phía trên của workfile này.

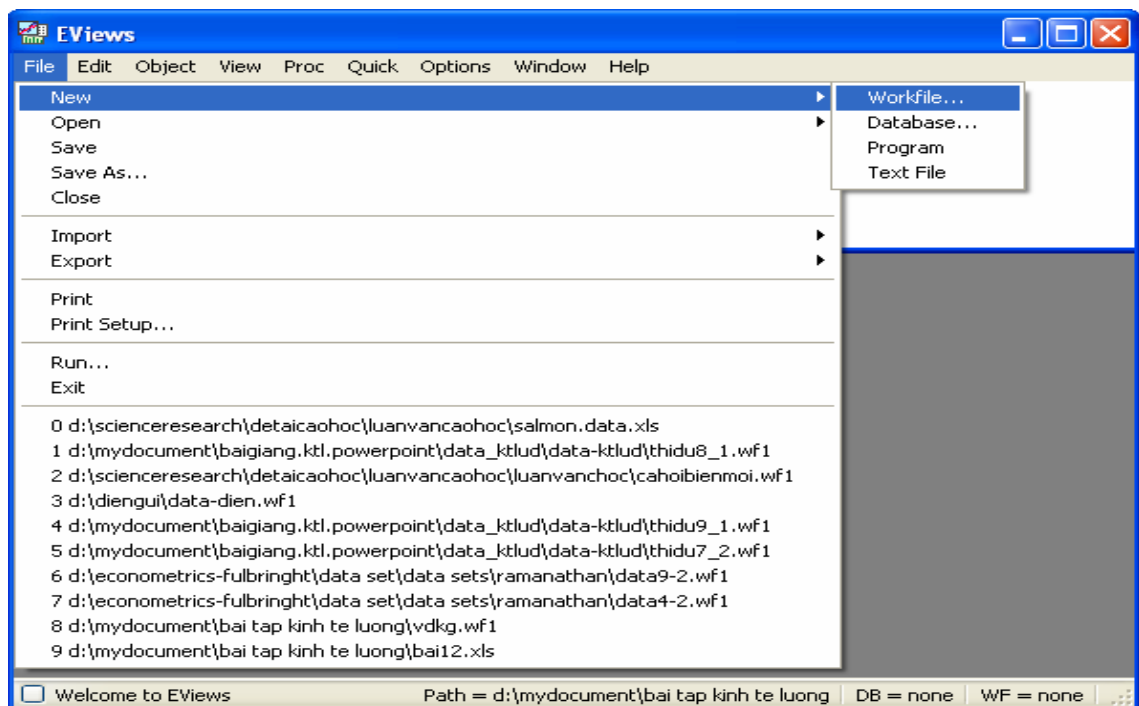
Bây giờ hãy nhấp **File**. Ghi nhớ rằng các workfile mới mở gần nhất sẽ hiện ra ở phía dưới cùng của menu trải ra phía dưới. Nếu các bạn muốn mở lại một trong số đó, chỉ việc nhấp vào tên của nó.



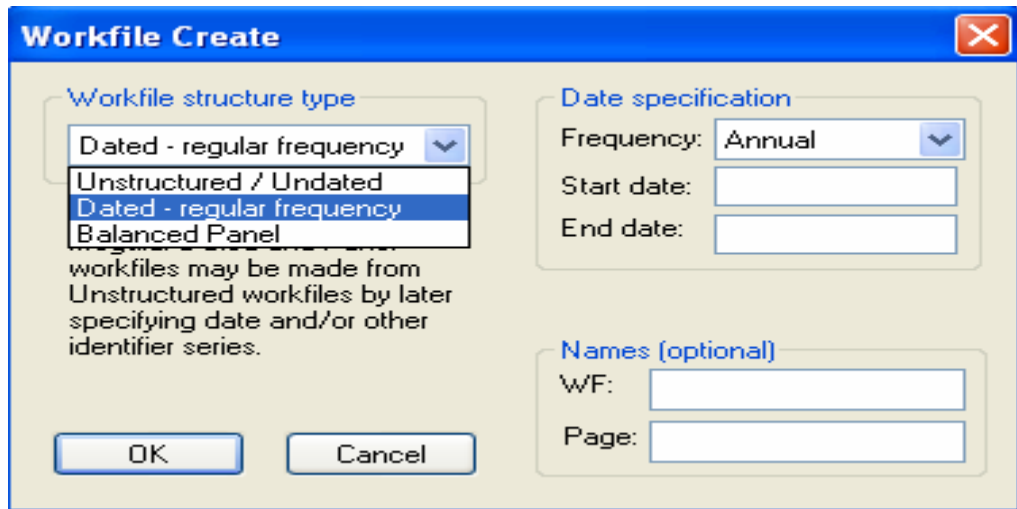
1.3. Tạo một file làm việc mới (Workfile) và nhập dữ liệu nghiên cứu.

1.3.1. Tạo file làm việc mới (New Workfile).

Sau khi đã khởi động EViews các bạn có thể tạo ra một file làm việc mới bằng cách nhấp chuột vào nút **File** trên menu chính, chọn **New/Workfile**, khi đó màn hình sẽ như sau:



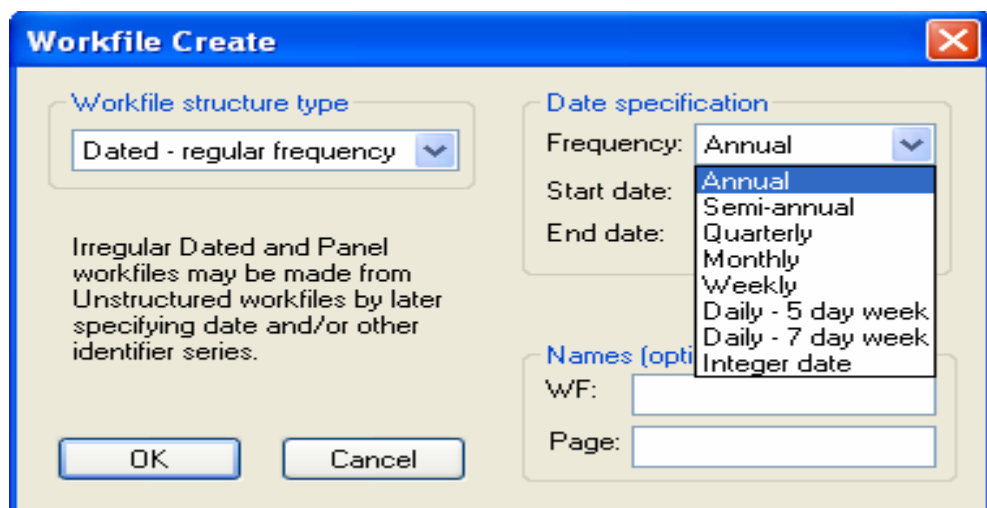
Nhấp chuột vào **Workfile**, màn hình sẽ xuất hiện cửa sổ **Workfile Create** như sau:



Nhấp chuột vào mũi tên ở khung **Workfile structure type** ta thấy có 3 loại cấu trúc file làm việc, đó là:

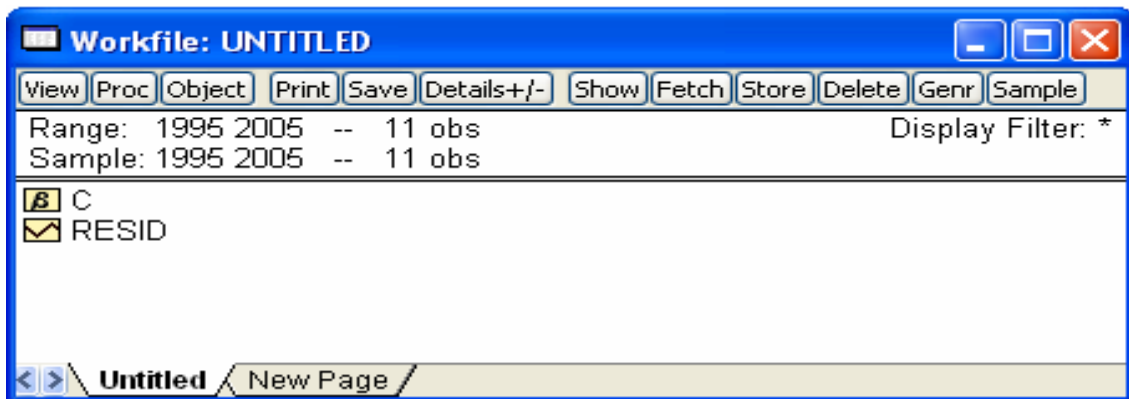
- **Date - regular frequency:** Cấu trúc file làm việc với dữ liệu theo thời gian.
- **Unstructure / Undated:** Cấu trúc file làm việc với dữ liệu chéo.
- **Balanced Panel:** Cấu trúc file làm việc với dữ liệu dạng bảng.

☛ Nếu dữ liệu của mẫu nghiên cứu là dữ liệu theo thời gian thì chọn **Date – regular frequency**. Tiếp theo, bạn nhấp chuột vào mũi tên ở khung **Date specification**, khi đó màn hình sẽ như sau:



Nhìn vào **frequency** ta chọn loại dữ liệu. Nếu là dữ liệu từng năm ta chọn **Annual**; nếu là dữ liệu nửa năm, chọn **Semi-annual**; nếu là dữ liệu từng quý, chọn **Quarterly**; nếu là dữ liệu từng tháng, chọn **Monthly**, nếu là dữ liệu tuần,

chọn **Weekly**,... Khi đã chọn loại dữ liệu thích hợp, trong khung **start date** ta gõ mốc thời gian đầu tiên (1995) - ví dụ: năm bắt đầu của mẫu dữ liệu nghiên cứu là **1995** và năm cuối cùng là **2005**, và trong khung **End date** ta gõ mốc thời gian cuối cùng (2005) của mẫu dữ liệu. Sau khi hoàn tất các thủ tục trên ta bấm **OK**, một file làm việc mới sẽ được tạo ra như sau:



• Nếu dữ liệu của mẫu nghiên cứu là dữ liệu chéo thì ta chọn **Unstructure / Undated**. Trong khung **Data range** ta gõ số quan sát của mẫu dữ liệu vào chỗ **Observations** và bấm **OK**. Ví dụ, cỡ mẫu là 55 ($n=55$) thì ta gõ số 55 vào. Khi đó, màn hình sẽ như sau:



• Nếu dữ liệu của mẫu nghiên cứu là dữ liệu dạng bảng thì ta chọn **Balanced Panel**.

*(Trong tài liệu này chủ yếu trình bày cách sử dụng Eviews với dữ liệu nghiên cứu theo thời gian và dữ liệu chéo, còn dữ liệu dạng bảng không được đề cập. Bạn đọc muốn biết cách sử dụng Eviews với dữ liệu dạng bảng có thể tìm đọc ở một số sách hướng dẫn khác hoặc xem phần **Help** trong Eviews).*

1.3.2. Nhập dữ liệu nghiên cứu.

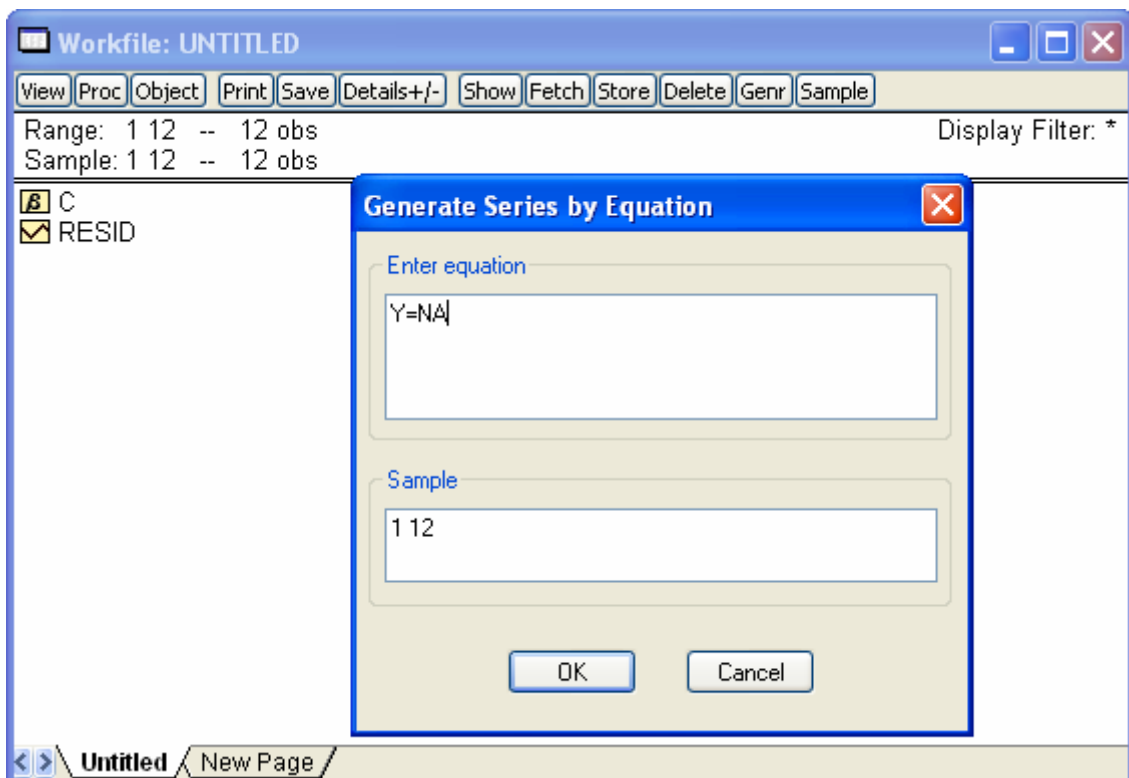
Sau khi đã tạo ra file làm việc mới, các bạn có thể nhập dữ liệu vào Eviews bằng nhiều cách khác nhau, sau đây là một số cách thức để nhập dữ liệu thường được sử dụng.

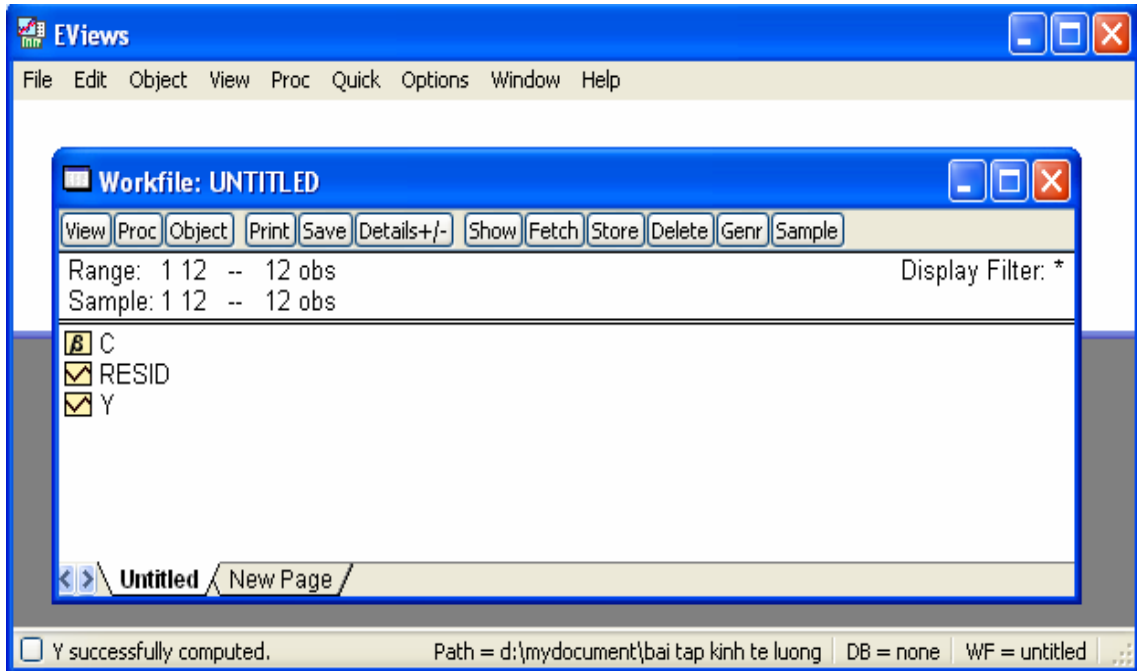
1.3.2.1. Nhập dữ liệu bằng cách copy và dán (Copy và Paste).

Sau khi các biến đã được xác định, các giá trị của chúng có thể được nhập vào bằng cách sử dụng lệnh copy, sau đó dán (Paste) vào màn hình của bảng tính EViews. Các bạn có thể copy và dán từ nhiều nguồn, bao gồm cả thư điện tử. Đôi khi các bạn phải lưu giữ dữ liệu này trước hết như là một tập tin ngôn ngữ (text file), sau đó copy và dán nó.

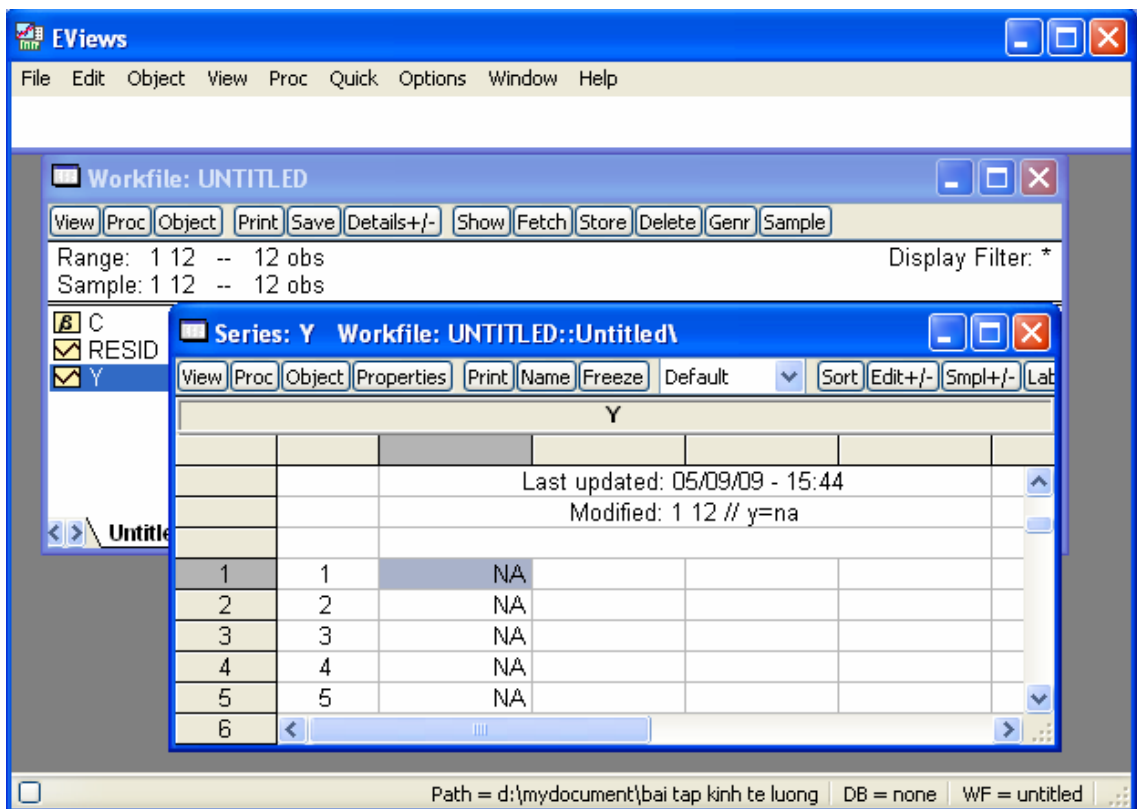
Trước khi bắt đầu thao tác **copy/paste** này, các bạn phải tạo ra các biến trong workfile mục tiêu. Hãy tưởng tượng rằng các bạn muốn copy và dán một biến có tên **Y** vào workfile của mình.

Đầu tiên, hãy mở workfile này, sau đó nhấp **Genr** và đánh máy **Y = NA**, sau đó bấm **OK**, động tác này tạo ra một biến có tên **Y** nhưng tất cả các giá trị của nó đều không có. Các bạn có thể quan sát ở các hình sau đây.



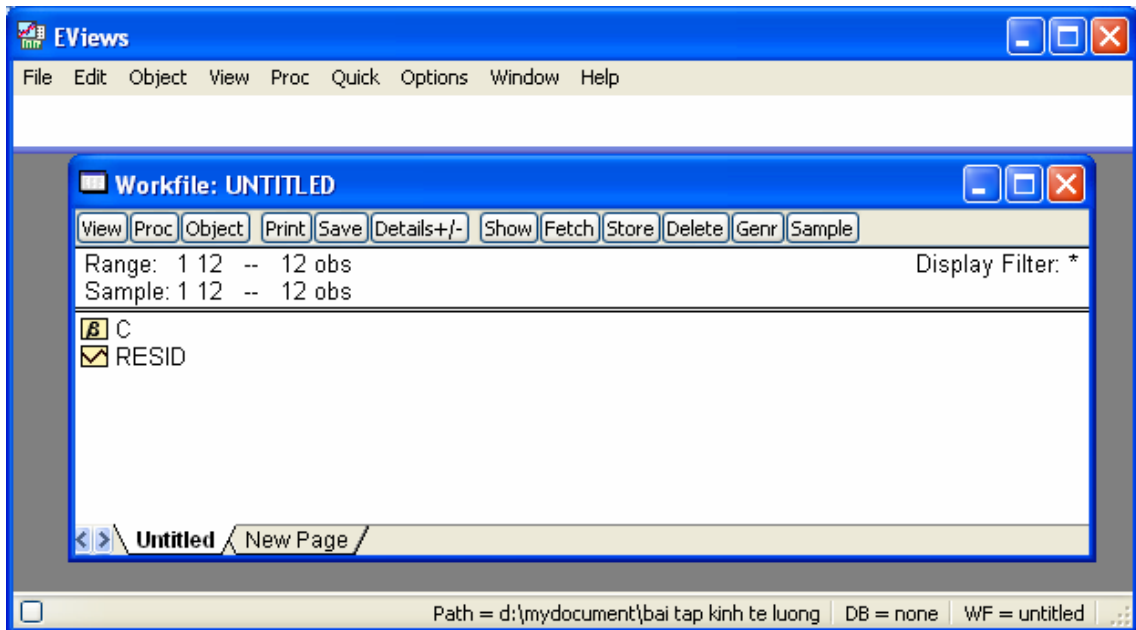


Mở chuỗi **Y**. Bây giờ nhấp **Edit+/-**. Lúc này, các bạn hãy bôi đen và copy dữ liệu cho **Y** từ nguồn của nó (chẳng hạn copy dữ liệu **Y** từ Excel), sau đó nhấp vào ô đầu tiên của chuỗi **Y** trên **EViews** và nhấp **Edit/Paste** ở phần menu chính (hoặc các bạn cũng có thể bấm chuột phải, sau đó **Paste**). Dữ liệu này rơi vào vị trí mà các bạn đã chọn. Hoàn tất thao tác này bằng cách tắt edit (nhấp **Edit+/-**).

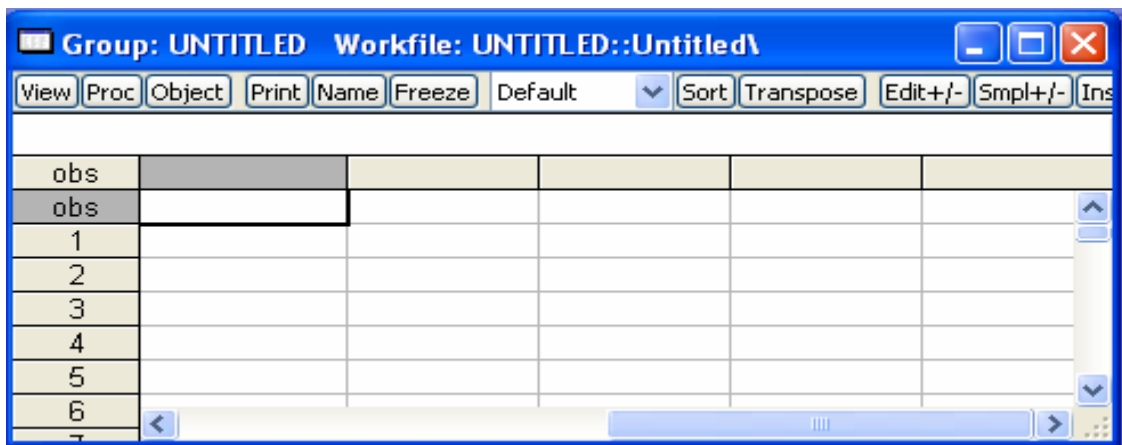


1.3.2.2. Nhập dữ liệu từ bàn phím.

Giả sử ta đã tạo ra một file làm việc mới có dạng như sau:



Từ cửa sổ **Eviews** ta chọn **Quick/Empty Group(Edit series)**, một cửa sổ sẽ xuất hiện như sau:

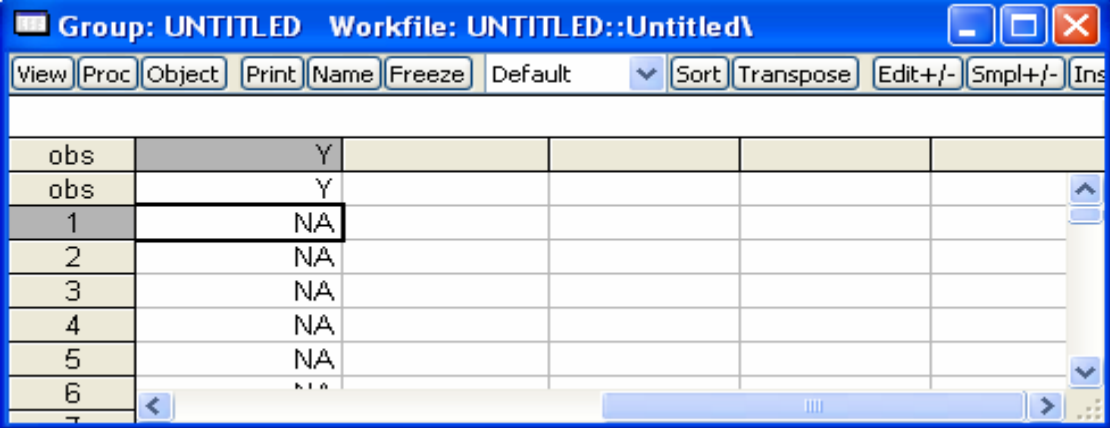


Cửa sổ này có cấu tạo giống như một bảng tính Excel, do đó chúng ta có thể nhập dữ liệu vào bảng tính này một cách dễ dàng như chúng ta nhập dữ liệu vào Excel vậy.

Nhìn vào cửa sổ trên ta thấy cột thứ nhất ghi thứ tự các quan sát. Từ cột thứ hai bạn có thể nhập dữ liệu của các biến, mỗi cột nhập vào một biến.

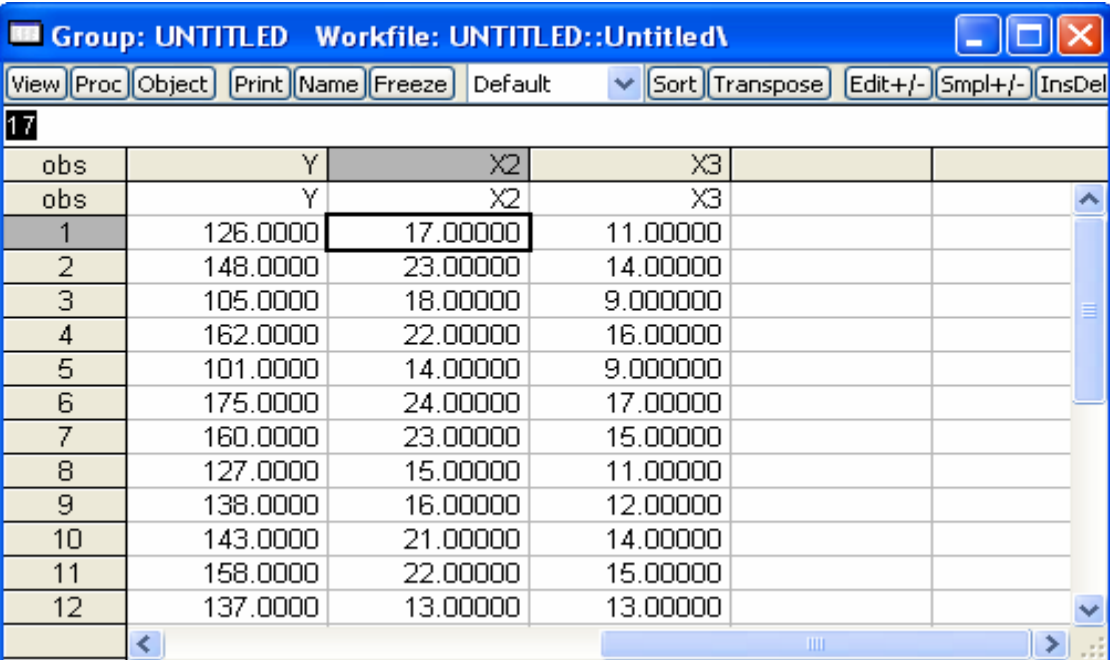
Ví dụ, để nhập dữ liệu của một mẫu quan sát ($n=12$) gồm các biến Y , X_2 , X_3 ta tiến hành như sau:

Đầu tiên ta nhập dữ liệu của biến Y trước (đĩ nhiên các bạn có thể nhập dữ liệu của biến X2 hoặc X3 trước đều được) vào cột thứ hai, ta nhấp chuột vào ô đầu tiên của cột này và gõ tên biến Y vào, sau đó bấm **Enter**. Màn hình sẽ có dạng như sau:



obs	Y			
obs	Y			
1	NA			
2	NA			
3	NA			
4	NA			
5	NA			
6	NA			

Lúc này các bạn gõ lần lượt giá trị quan sát của biến Y vào ô đóng khung (bên trong có chữ NA – Not Available). Sau khi nhập xong dữ liệu của biến Y, các bạn có thể tiến hành nhập dữ liệu của các biến X2, X3 một cách tương tự. Kết quả nhập liệu của các biến sau khi hoàn thành sẽ được thể hiện như màn hình sau:



obs	Y	X2	X3	
obs	Y	X2	X3	
1	126.0000	17.00000	11.00000	
2	148.0000	23.00000	14.00000	
3	105.0000	18.00000	9.000000	
4	162.0000	22.00000	16.00000	
5	101.0000	14.00000	9.000000	
6	175.0000	24.00000	17.00000	
7	160.0000	23.00000	15.00000	
8	127.0000	15.00000	11.00000	
9	138.0000	16.00000	12.00000	
10	143.0000	21.00000	14.00000	
11	158.0000	22.00000	15.00000	
12	137.0000	13.00000	13.00000	

Sau khi nhập liệu xong, đóng cửa sổ dữ liệu bằng cách nhấp chuột vào ô có dấu X ở góc trên bên phải cửa sổ này.

1.3.2.3. Nhập dữ liệu từ các phần mềm khác.

Dữ liệu có thể được nhập vào từ các tập tin Lotus, Excel, SPSS, MINITAB hoặc ASCII.

Để nhập được một file số liệu từ Excel vào Eviews bạn cần phải ghi nhớ:

- Tần suất (loại) của số liệu trong file này là gì? (Số liệu chéo hay số liệu theo thời gian. Nếu là số liệu theo thời gian thì tần suất thời gian là tháng, quý, năm, tuần hay ngày,...).

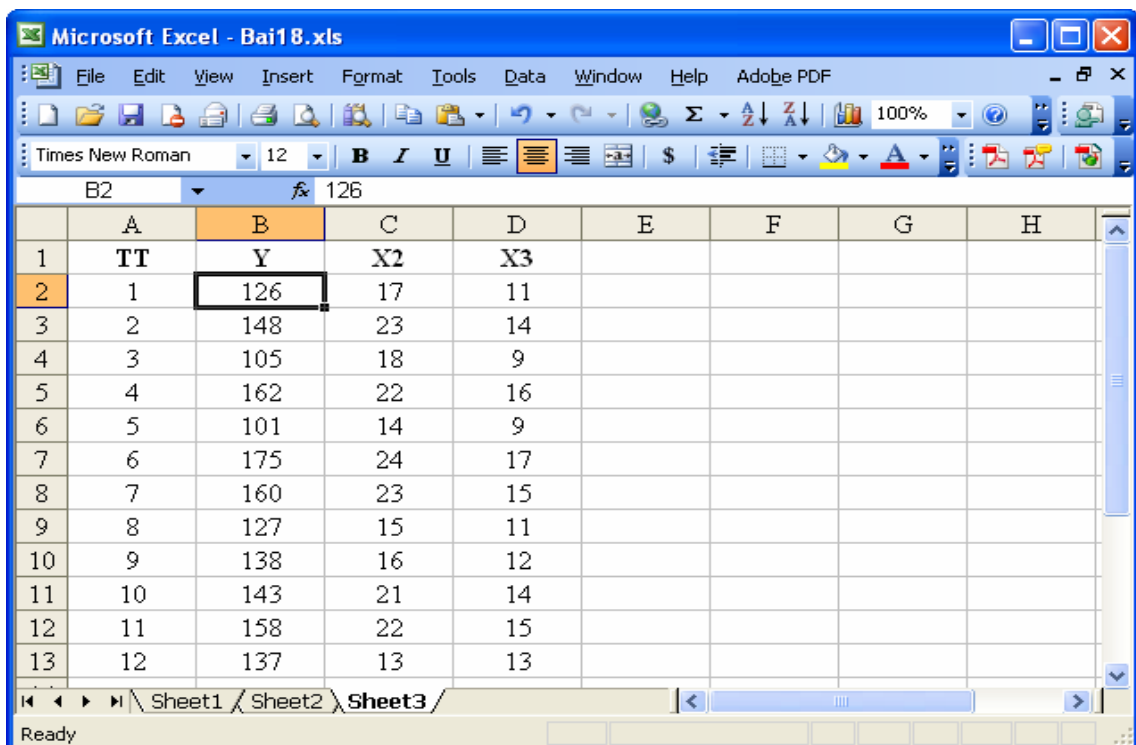
- Thời điểm bắt đầu và thời điểm kết thúc đối với chuỗi thời gian, nếu là dữ liệu chéo thì có bao nhiêu quan sát tất cả (cỡ mẫu là bao nhiêu).

- Có bao nhiêu biến số và tên của các biến số đó. Bạn cũng cần ghi nhớ là các biến trong file Excel là ghi theo cột hay theo hàng.

- Số liệu bắt đầu được ghi ở ô nào, có tên biến số ở ô đầu tiên hay không?

- Số liệu ghi trong file Excel được ghi ở sheet nào? (Tên sheet?).

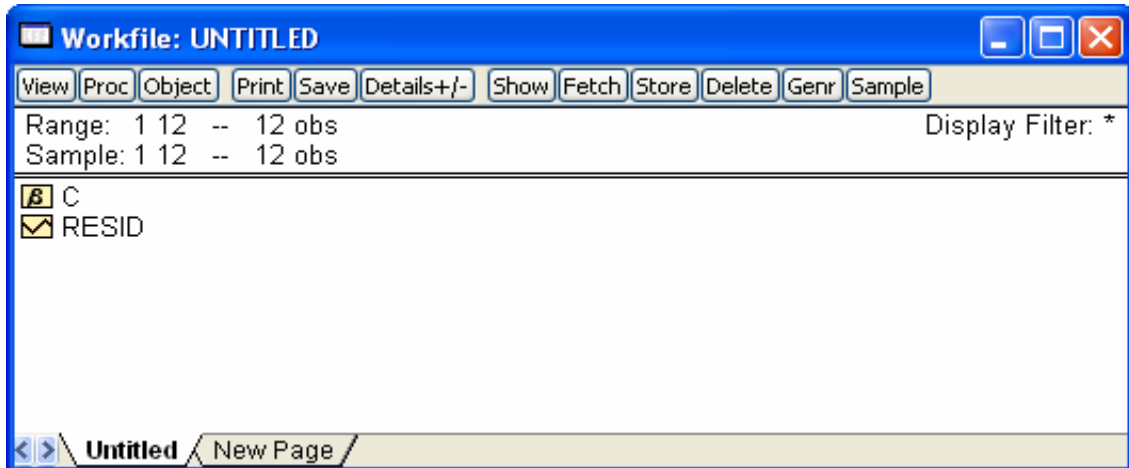
Chúng ta biểu diễn với một tập tin Excel có tên **Bai18.xls** chứa dữ liệu của ba biến gồm: Y (Doanh số), X2 (Chi phí quảng cáo), X3 (Tiền lương của nhân viên tiếp thị) của 12 công ty. Dữ liệu trong file Excel khi mở ra màn hình của nó có dạng như sau:



	A	B	C	D	E	F	G	H
1	TT	Y	X2	X3				
2	1	126	17	11				
3	2	148	23	14				
4	3	105	18	9				
5	4	162	22	16				
6	5	101	14	9				
7	6	175	24	17				
8	7	160	23	15				
9	8	127	15	11				
10	9	138	16	12				
11	10	143	21	14				
12	11	158	22	15				
13	12	137	13	13				

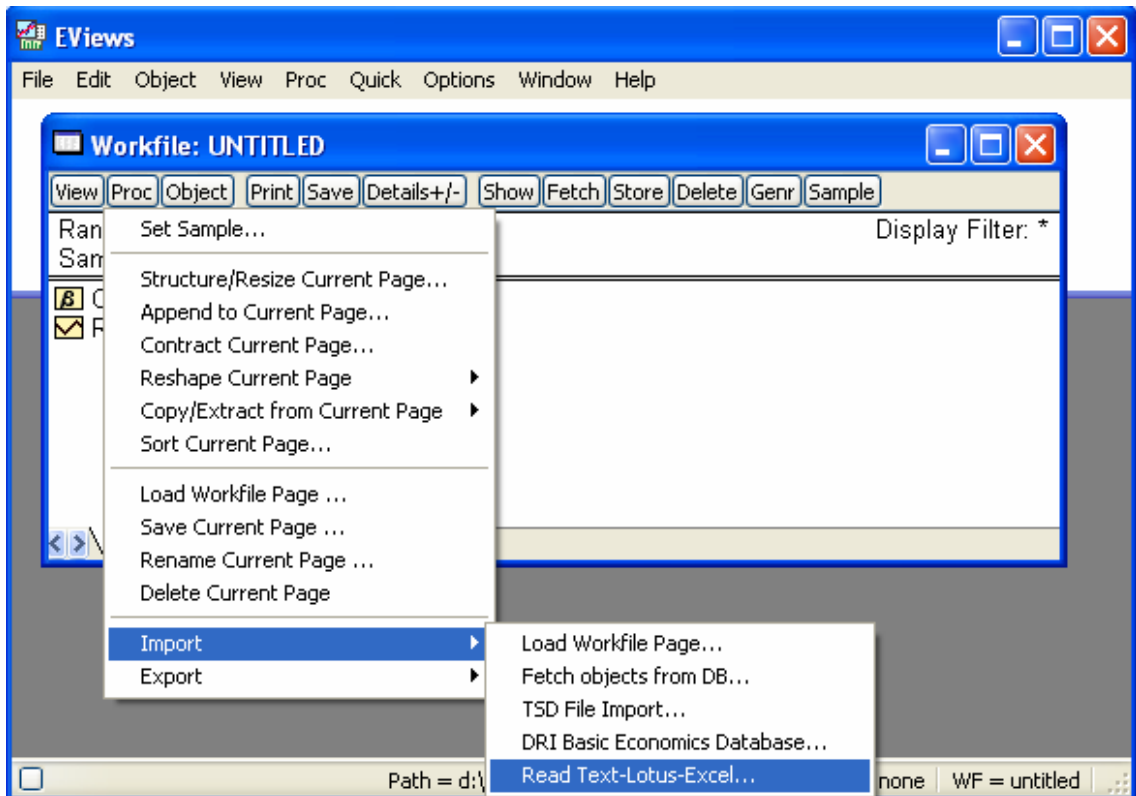
Quan sát file dữ liệu này ta thấy dữ liệu nghiên cứu là dữ liệu chéo; có tổng cộng là 12 quan sát; có ba biến số là Y, X2, X3; cả ba biến số này đều được ghi theo cột; số liệu bắt đầu được ghi ở ô **B2**, với tên biến số là **Y** ở ô đầu tiên (B1); và tập dữ liệu này được ghi ở **Sheet3**. Các bước tiến hành như sau:

Đầu tiên các bạn phải tạo ra một file làm việc (Workfile) mới. Cách tạo ra một file làm việc đã được trình bày ở phần trên. File làm việc sau khi được tạo ra sẽ có màn hình như sau:

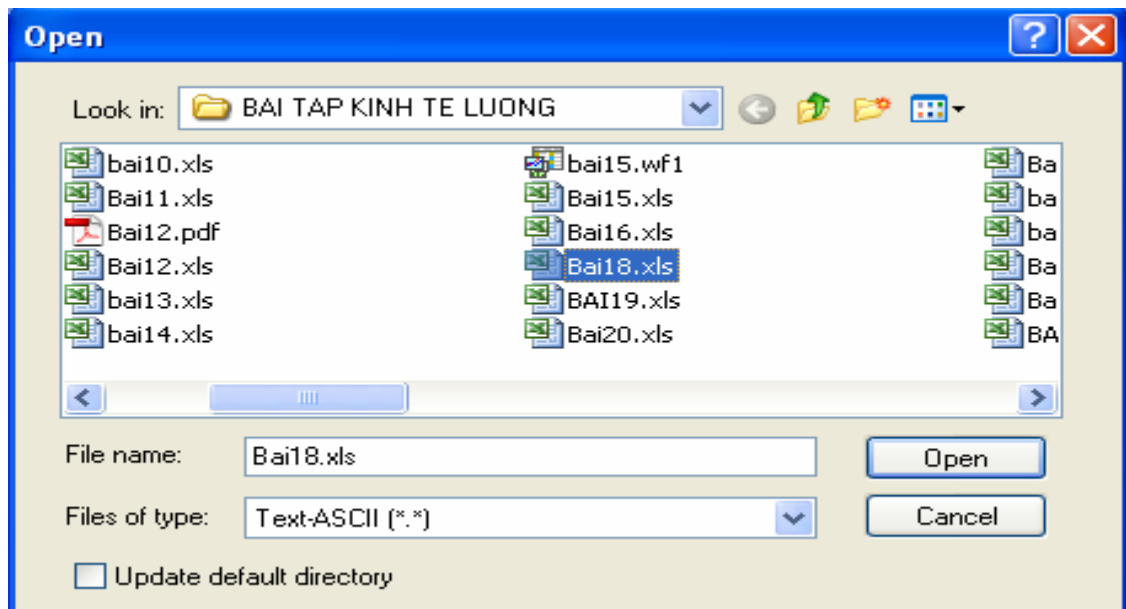


Trong cửa sổ này ta thấy có vector hệ số “C” và vector phần dư “Resid” được Eviews tự động tạo ra. Vector hệ số “C” chứa các hệ số hồi quy được ước lượng. Vector phần dư “Resid” chứa phần dư. Do vậy, không được dùng “C” và “Resid” để đặt tên cho các biến khác.

Từ cửa sổ Workfile này bạn nhấp chuột vào nút **Proc/Import/Read Text-Lotus-Excel...** Bạn sẽ được nhìn thấy như ở màn hình sau:



Sau khi chọn **Proc/Import/Read Text-Lotus-Excel...** như chỉ dẫn ở hình trên, và bạn phải chỉ ra đường dẫn và tên file. Trong ví dụ này file **Bai18.xls** có đường dẫn là: **D:\MYDOCUMENT\BAI TAP KINH TE LUONG\ Bai18.xls**. Khi đó màn hình được mở ra như sau:



Sau khi bạn nhấp chuột vào nút **“Open”**, cửa sổ dưới đây sẽ xuất hiện.

The image shows a dialog box titled "Excel Spreadsheet Import". It contains several sections:

- Data order:** Two radio buttons. "By Observation - series in columns" is selected. "By Series - series in rows" is unselected.
- Upper-left data cell:** A text box containing "B2".
- Excel 5+ sheet name:** A text box containing "SHEET3".
- Names for series or Number if named in file:** A text box containing "3".
- Write date/obs:** A checkbox that is unchecked.
- EViews date format:** Three radio buttons: "EViews date format", "First calendar day", and "Last calendar day". All are unselected.
- Write series names:** A checkbox that is unchecked.
- Import sample:** A text box containing "1 12".
- Reset sample to:** Three checkboxes: "Current sample", "Workfile range", and "To end of range". All are unchecked.
- Buttons:** "OK" and "Cancel" buttons at the bottom right.

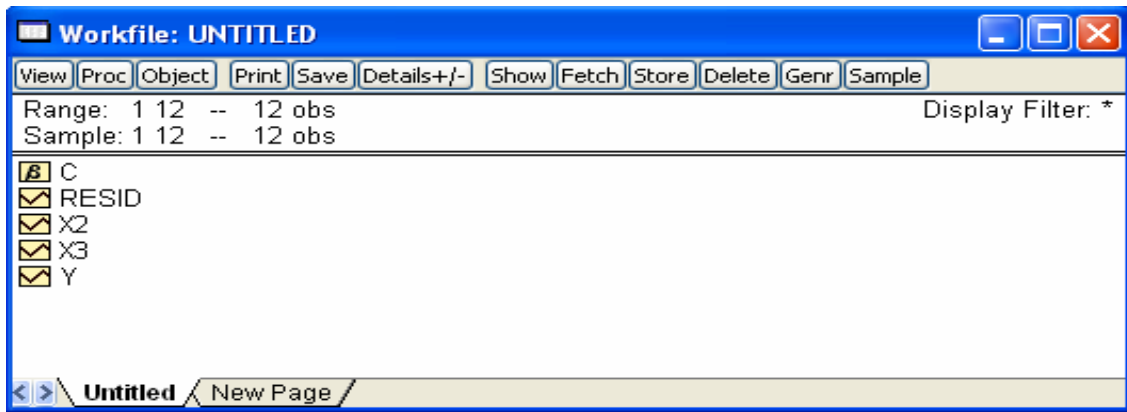
Trong khung “**Data order**” máy tính đã mặc định dữ liệu được ghi theo cột. Vì vậy, ta không cần phải chọn nữa vì dữ liệu của file **Bai18.xls** được ghi theo cột. Nếu dữ liệu ghi theo hàng ta chọn “**By Series – series in rows**”.

Trong khung “**Upper - left data cell**” máy tính đã mặc định dữ liệu được ghi bắt đầu ở vị trí ô **B2**. Do đó, ta không cần thay đổi gì cả vì dữ liệu của file **Bai18.xls** được ghi bắt đầu ở ô B2. Giả sử dữ liệu bắt đầu ở ô C2 hay A2 thì ta phải thay B2 bằng C2 hay A2 cho phù hợp.

Trong khung “**Excel 5 + sheet name**” chúng ta phải gõ vào chữ “**sheet3**” vì dữ liệu của file **Bai18.xls** được ghi ở sheet3. Nếu bạn không gõ vào tên sheet (**sheet name**) là gì, khi đó máy tính sẽ mặc nhiên lấy từ sheet1.

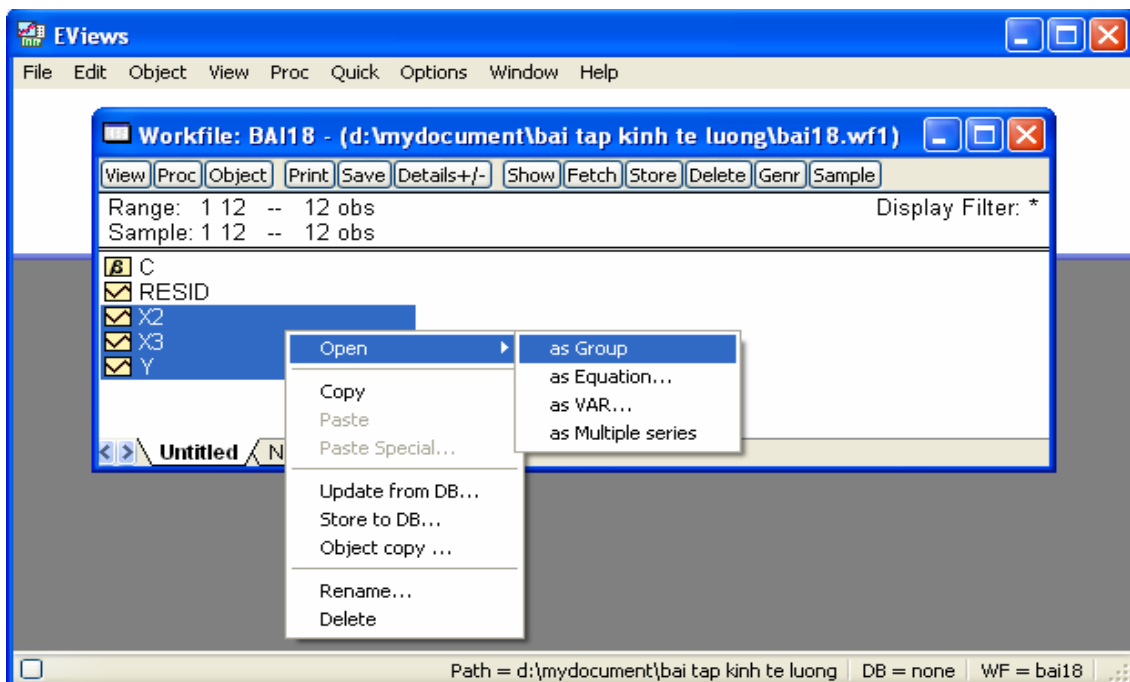
Trong khung “**Names for series or...**” ta nhập vào tên của các biến số, mỗi biến số cách nhau ít nhất một khoảng trắng. Trong trường hợp file có nhiều biến số thì ta chỉ cần nhập số biến số. Trong ví dụ này ta nhập số biến số là **3**.

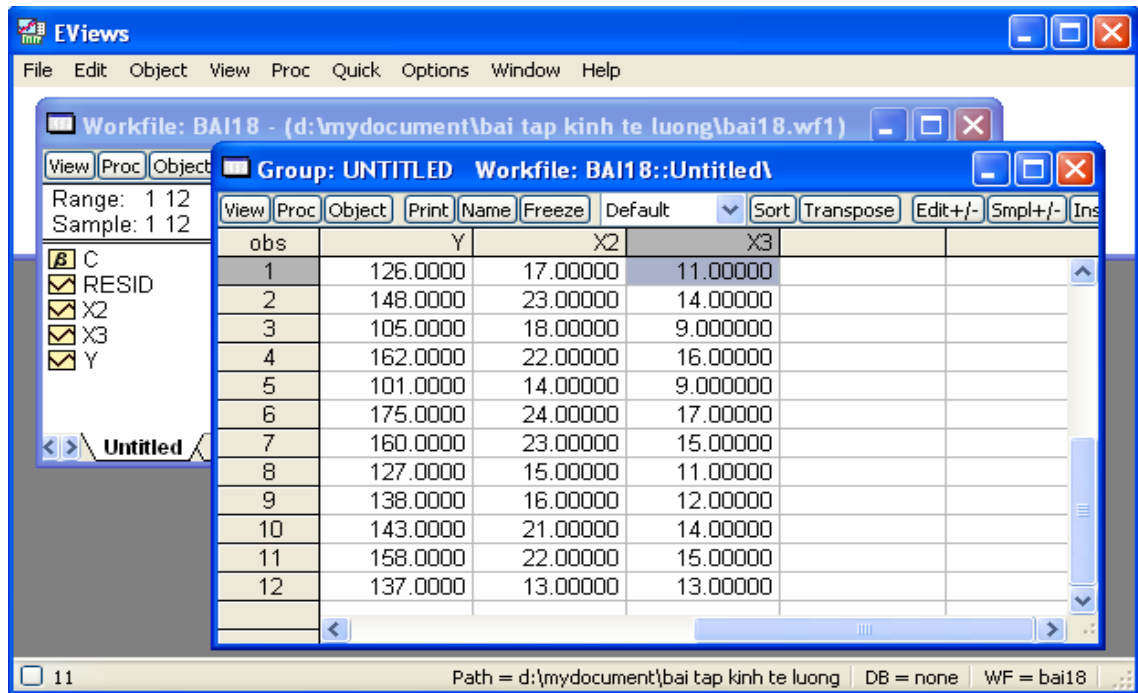
Sau khi làm xong các bước ở trên, bạn chọn “**OK**” để hoàn tất việc nhập liệu. Khi đó, một cửa sổ sẽ xuất hiện như sau:



Workfile này chưa có tên. Bạn có thể lưu workfile này bằng cách nhấp vào “**Save**”, đặt một tên mà bạn muốn, lưu ở thư mục nào, ổ đĩa nào là tùy bạn. Ở đây tôi lưu với tên là **Bai18.wf1**. Lưu ý, phần mở rộng của file Eviews là **wf1** máy sẽ tự gán. Một khi bạn đã lưu lại rồi, sau này bạn muốn sử dụng workfile này bạn không cần phải nhập dữ liệu lại từ đầu.

Bằng cách nhấp đúp chuột vào từng biến, các bạn có thể kiểm tra xem dữ liệu đã được nhập vào như mong muốn hay chưa. Các bạn cũng có thể mở nhiều biến cùng lúc trên một nhóm (Group) theo thứ tự mà mình muốn bằng cách bạn nhấn phím **Ctrl** và nhấp chuột vào biến số mong muốn. Sau khi đã chọn xong các biến số, bạn chỉ cần đưa chuột vào bất kỳ vị trí nào trong phần đã được đánh dấu rồi bấm chuột phải, chọn **Open/as Group**. Các bạn có thể nhìn thấy ở các hình sau.





The screenshot shows the EViews software interface. The main window displays a data table with the following columns: 'obs', 'Y', 'X2', and 'X3'. The data is as follows:

obs	Y	X2	X3
1	126.0000	17.00000	11.00000
2	148.0000	23.00000	14.00000
3	105.0000	18.00000	9.000000
4	162.0000	22.00000	16.00000
5	101.0000	14.00000	9.000000
6	175.0000	24.00000	17.00000
7	160.0000	23.00000	15.00000
8	127.0000	15.00000	11.00000
9	138.0000	16.00000	12.00000
10	143.0000	21.00000	14.00000
11	158.0000	22.00000	15.00000
12	137.0000	13.00000	13.00000

The interface also shows a menu bar (File, Edit, Object, View, Proc, Quick, Options, Window, Help) and a toolbar with various options like View, Proc, Object, Print, Name, Freeze, Default, Sort, Transpose, Edit+/-, Smpl+/-, and Ins. The status bar at the bottom indicates the path, database name, and workfile name.

Như vậy, ta đã nhập xong dữ liệu từ một file Excel vào Eviews theo mong muốn. Trong nghiên cứu người ta thường thích nhập liệu bằng cách này hơn.

Chương 2: LÀM VIỆC VỚI MỘT WORKFILE

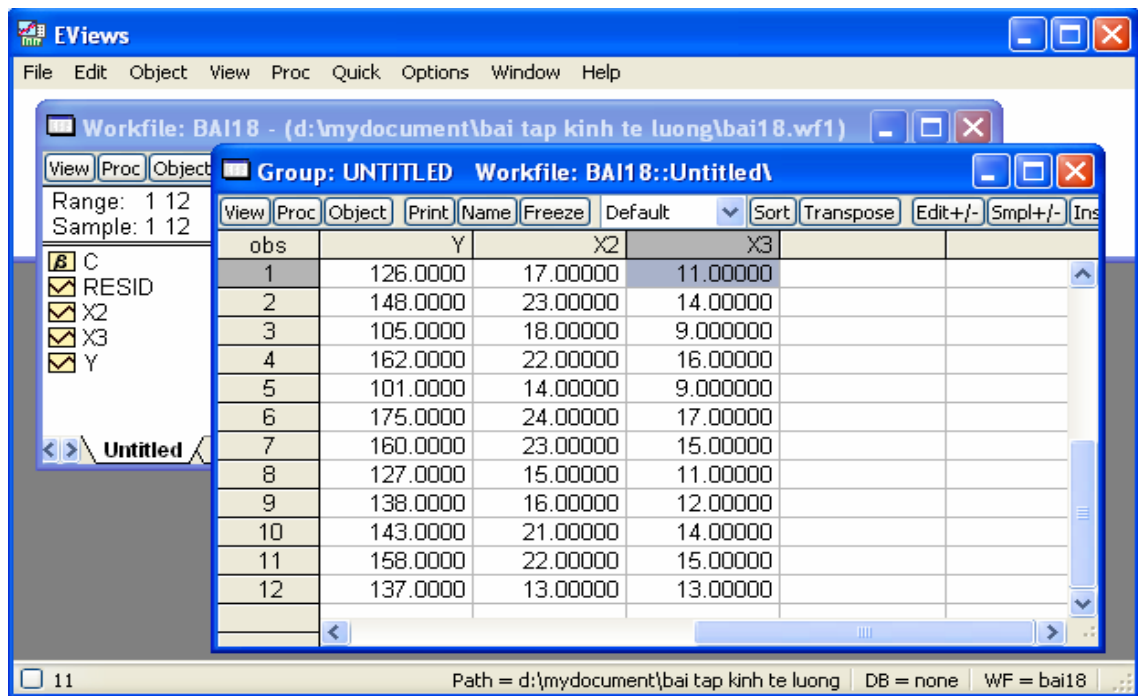
Sau khi đã tạo ra một workfile mới và nhập dữ liệu xong, chúng ta có thể tiến hành các phân tích như: Tính các trị thống kê mô tả, vẽ đồ thị, ước lượng hàm kinh tế lượng, kiểm định giả thuyết, dự báo...

2.1. Lựa chọn các biến từ một Workfile.

Sau khi khởi động Eviews, bạn hãy mở một workfile đã được tạo ra từ trước bằng cách nhấp vào **File/Open/Workfile**, sau đó tìm đến file cần mở (ví dụ ở đây là file có tên **Bai18.wf1**) hãy nhấp đúp vào **Bai18.wf1**.

Với mục đích lựa chọn một biến (hoặc nhiều biến) để làm việc, bạn hãy bôi đen chúng. Nếu bạn muốn bôi đen nhiều biến, thì hãy giữ phím **Ctrl** và nhấp vào mỗi biến, hoặc kéo con trỏ qua các biến mà bạn muốn bôi đen. Hãy bôi đen **Y, X2** và **X3**.

Nhấp đúp vào một trong các biến đã bôi đen (**hoặc nhấp chuột phải**), chọn **Open/ as Group**, và sau đó cả ba biến đều hiện trên bảng tính như hình sau:

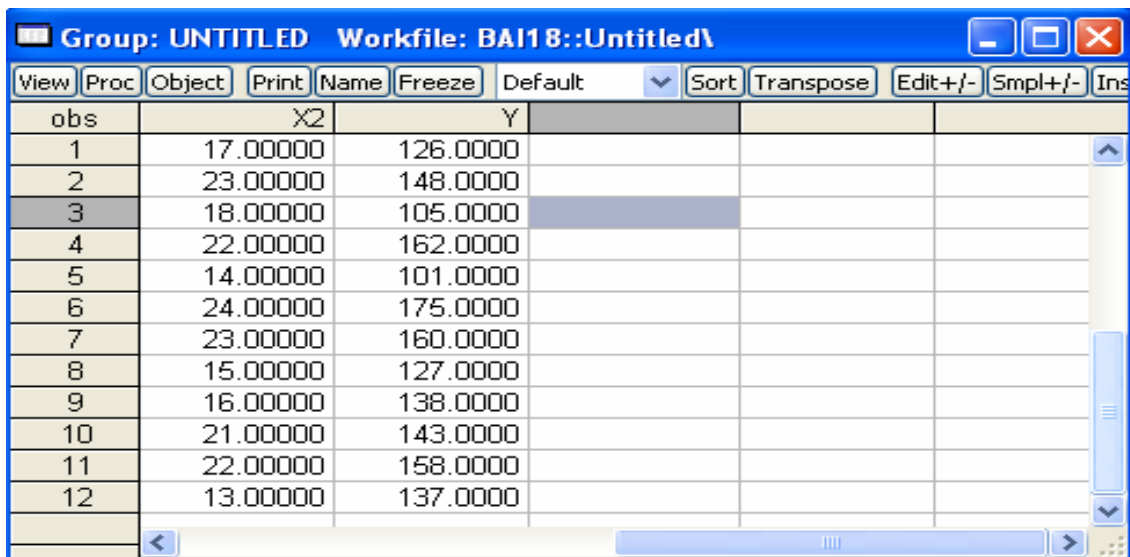


obs	Y	X2	X3
1	126.0000	17.00000	11.00000
2	148.0000	23.00000	14.00000
3	105.0000	18.00000	9.000000
4	162.0000	22.00000	16.00000
5	101.0000	14.00000	9.000000
6	175.0000	24.00000	17.00000
7	160.0000	23.00000	15.00000
8	127.0000	15.00000	11.00000
9	138.0000	16.00000	12.00000
10	143.0000	21.00000	14.00000
11	158.0000	22.00000	15.00000
12	137.0000	13.00000	13.00000

Nhập **name** và đặt tên cho nhóm **Y, X2, X3** với tên nhóm mà bạn thích. Bạn có thể kiểm tra chuỗi Y, X2, X3 bằng cách sử dụng các phím $\leftarrow \uparrow \rightarrow \downarrow$ để di chuyển trên bảng tính.

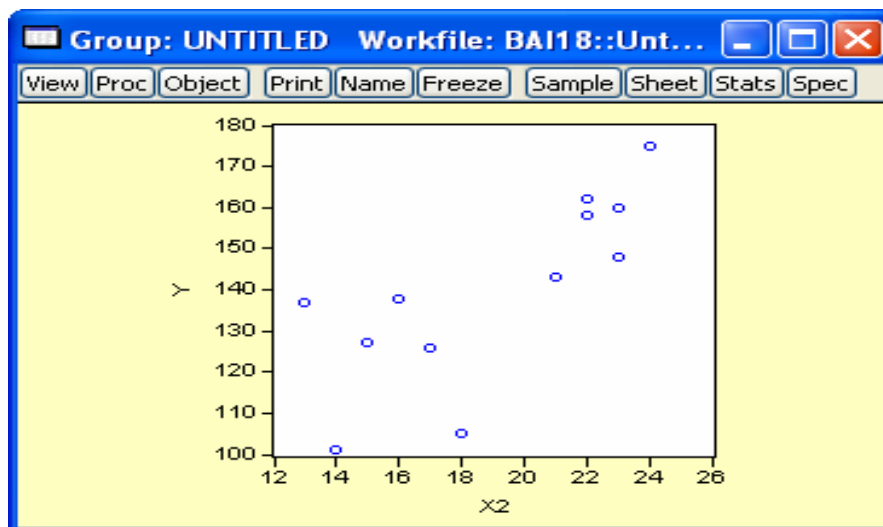
Lưu ý: Nếu bạn chỉ vẽ đồ thị cho một biến thì nên mở một biến thôi. Nếu bạn chọn **Graph** thì máy sẽ vẽ đồ thị của tất cả các biến số có trong **Group** trên cùng một hệ trục tọa độ, với trục tung là biến số được chọn, trục hoành là thứ tự các quan sát (nếu là dữ liệu chéo) hay thời gian (nếu là dữ liệu theo thời gian). Các biến số được vẽ với các màu khác nhau để bạn dễ nhận biết. Nếu bạn chọn **Multiple Graphs** thì mỗi biến số được vẽ trên một đồ thị riêng biệt.

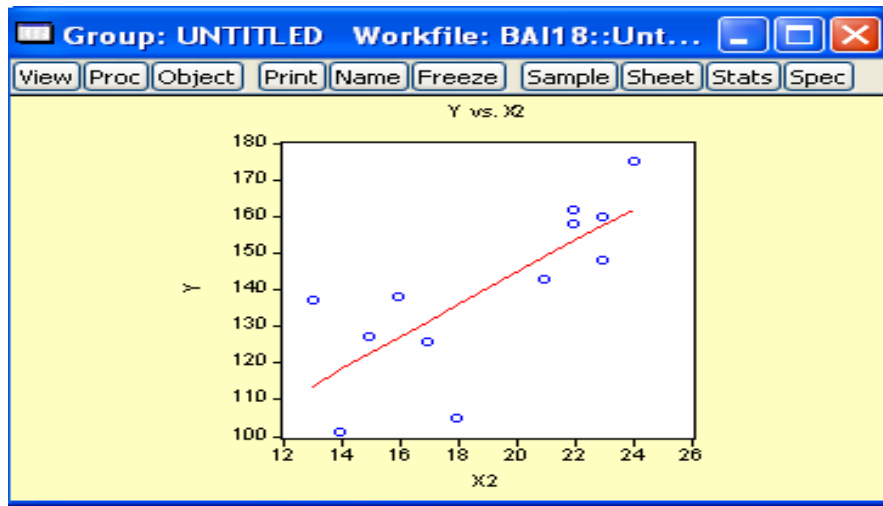
• Nếu bạn muốn vẽ đồ thị phân tán của hai biến nào đó (ví dụ, ta vẽ đồ thị phân tán của Y theo X2). Để vẽ đồ thị này trước tiên bạn hãy mở biến X2 và Y trên cùng một nhóm (ghi nhớ là biến số nằm trên trục hoành chọn trước rồi mới chọn biến số trên trục tung), xem hình sau:



obs	X2	Y
1	17.00000	126.0000
2	23.00000	148.0000
3	18.00000	105.0000
4	22.00000	162.0000
5	14.00000	101.0000
6	24.00000	175.0000
7	23.00000	160.0000
8	15.00000	127.0000
9	16.00000	138.0000
10	21.00000	143.0000
11	22.00000	158.0000
12	13.00000	137.0000

Từ cửa sổ này bạn chọn **View/Graph/Scatter/simple Scatter** hoặc **View/Graph/Scatter/ Scatter with Regression,...** khi đó đồ thị phân tán sẽ như sau:

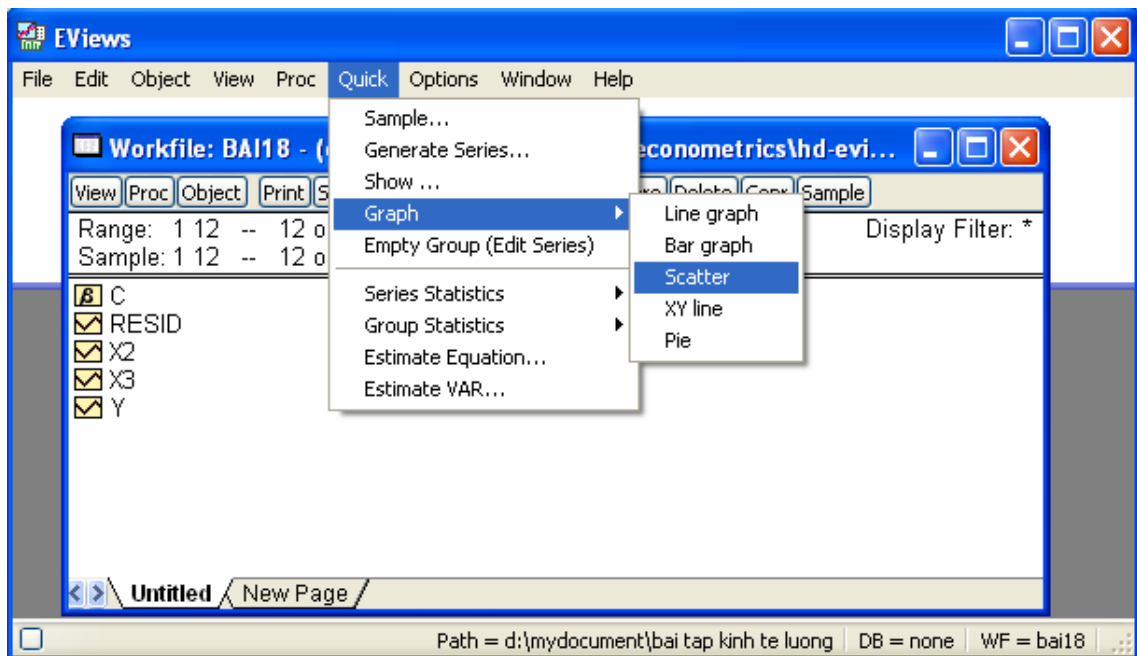




Bạn có thể thấy sự khác nhau của hai cách chọn với hai đồ thị khác nhau như hình trên. Trên cửa sổ Group có nhiều thực đơn khác nhau, mỗi thực đơn có nhiều lựa chọn khác nhau, ở đây không trình bày hết các lựa chọn đó. Bạn đọc có thể thử từng lựa chọn một rồi xem kết quả để rút ra bài học riêng cho mình.

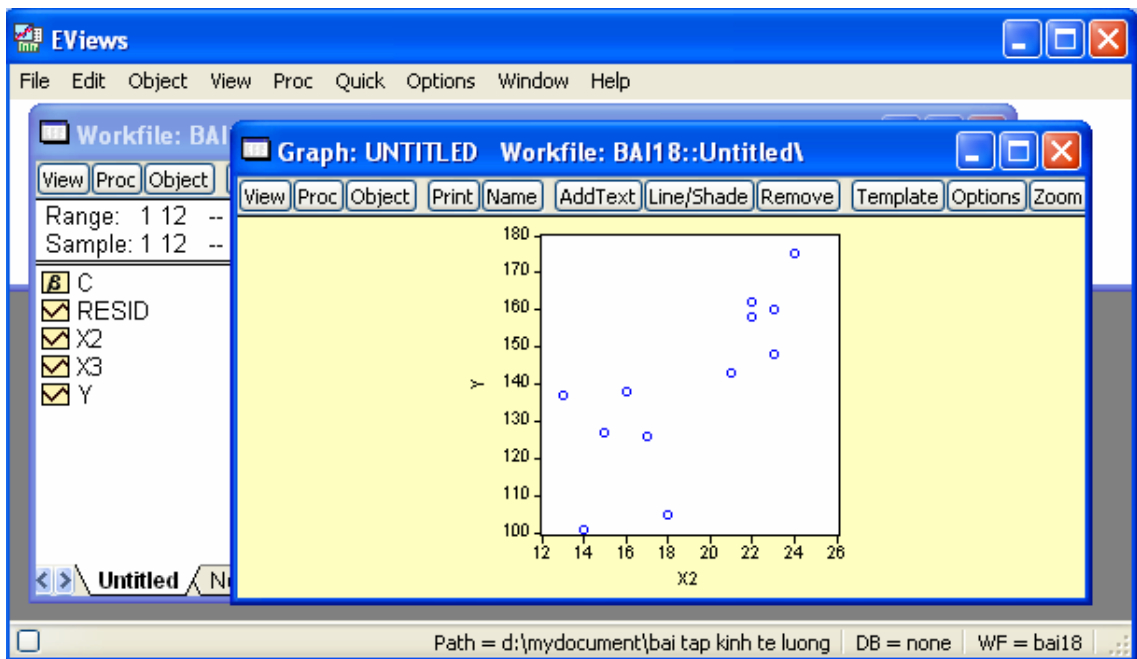
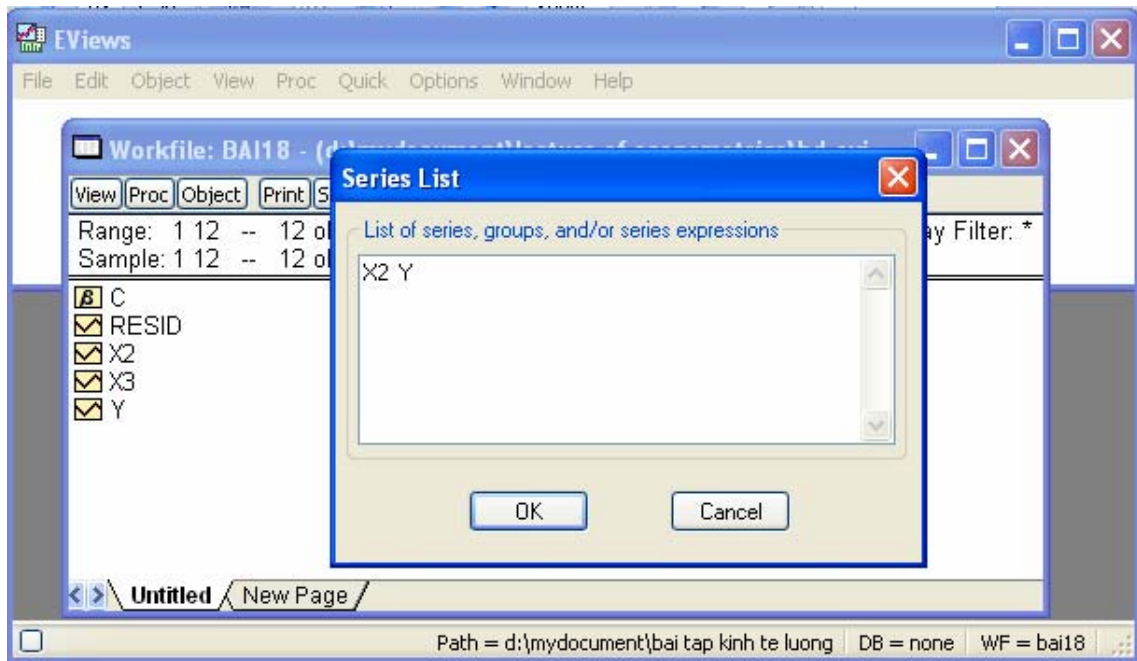
Vẽ đồ thị không cần mở chuỗi dưới dạng bảng.

Bây giờ hãy trở lại workfile. Hãy vào menu chính và nhấp vào **Quick/Graph** dùng để vẽ đồ thị linh động hơn. Cách vẽ này không cần phải mở chuỗi dữ liệu như cách đã vẽ ở trên.



Một cửa sổ mở ra yêu cầu bạn xác định chuỗi hoặc các nhóm, bạn cần đánh máy **X2 Y**, mỗi biến cách nhau một khoảng trắng và nhấn **OK**. Biến mà bạn

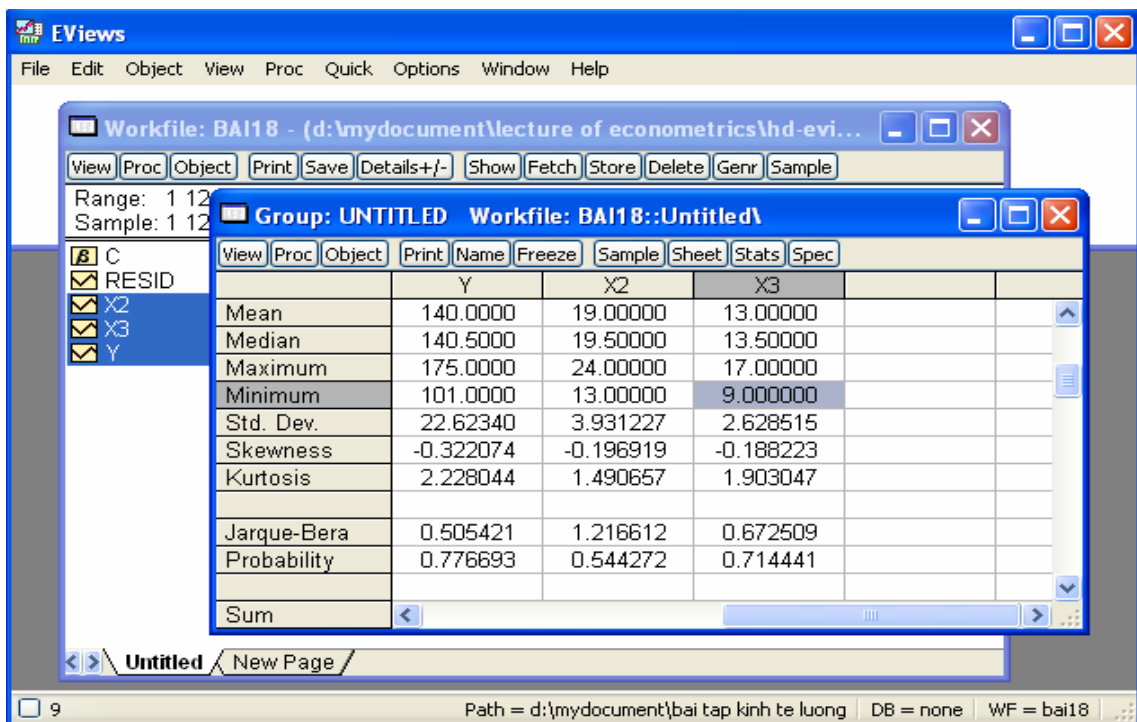
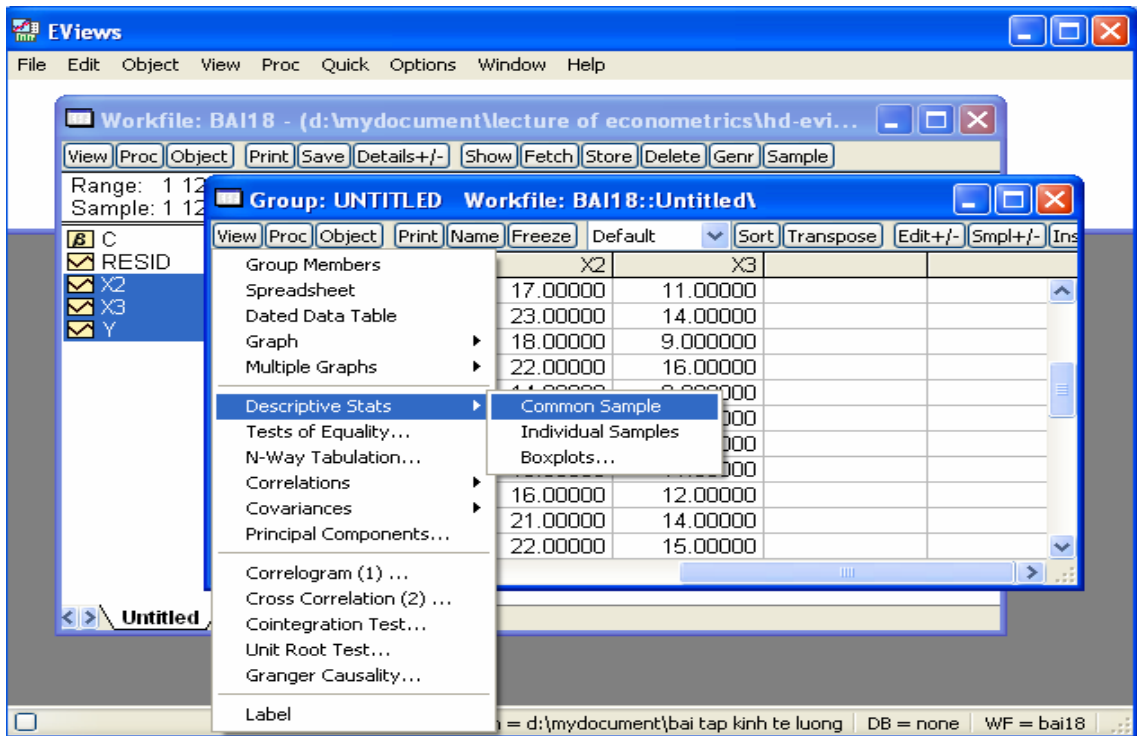
liệt kê đầu tiên được vẽ trên trục hoành. Hãy lưu ý tới sự đa dạng của các lựa chọn trong menu.



2.3. Tính các trị thống kê mô tả của mẫu.

Giả sử ta muốn tính các trị thống kê mô tả của các biến Y, X2, và X3 ta làm như sau:

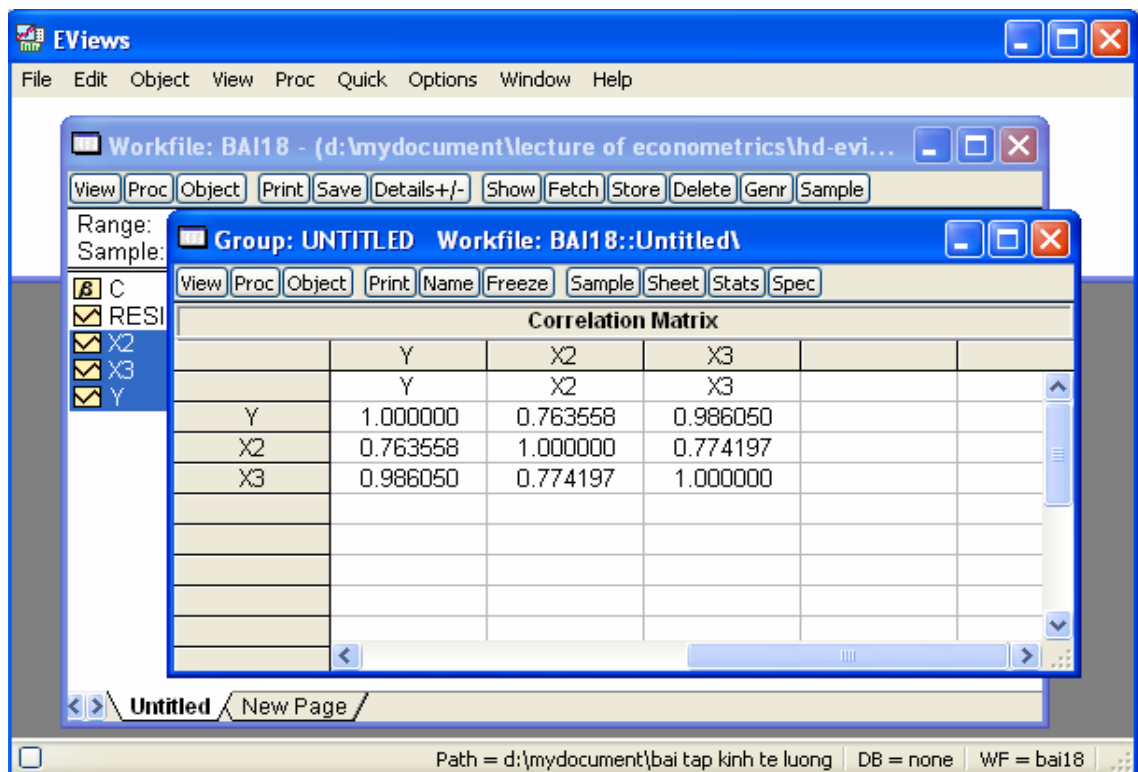
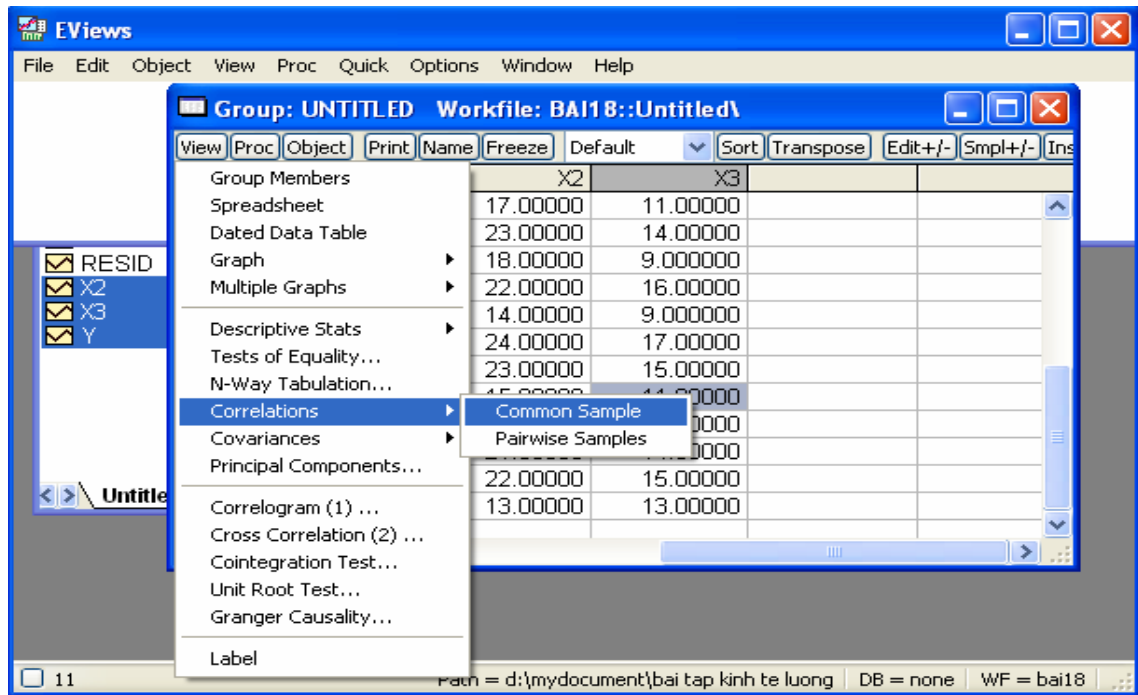
Trước tiên các bạn lựa chọn và mở các biến cần tính trên một Group, sau đó chọn **View/Descriptive Statistics/Common Sample**.



2.4. Tính hệ số tương quan (ma trận tương quan) giữa các biến.

Giả sử ta muốn tìm ma trận tương quan của các biến Y, X2, và X3 ta làm như sau:

Trước tiên các bạn lựa chọn và mở các biến cần tính trên một Group, sau đó chọn **View/Correlations/Common Sample**.

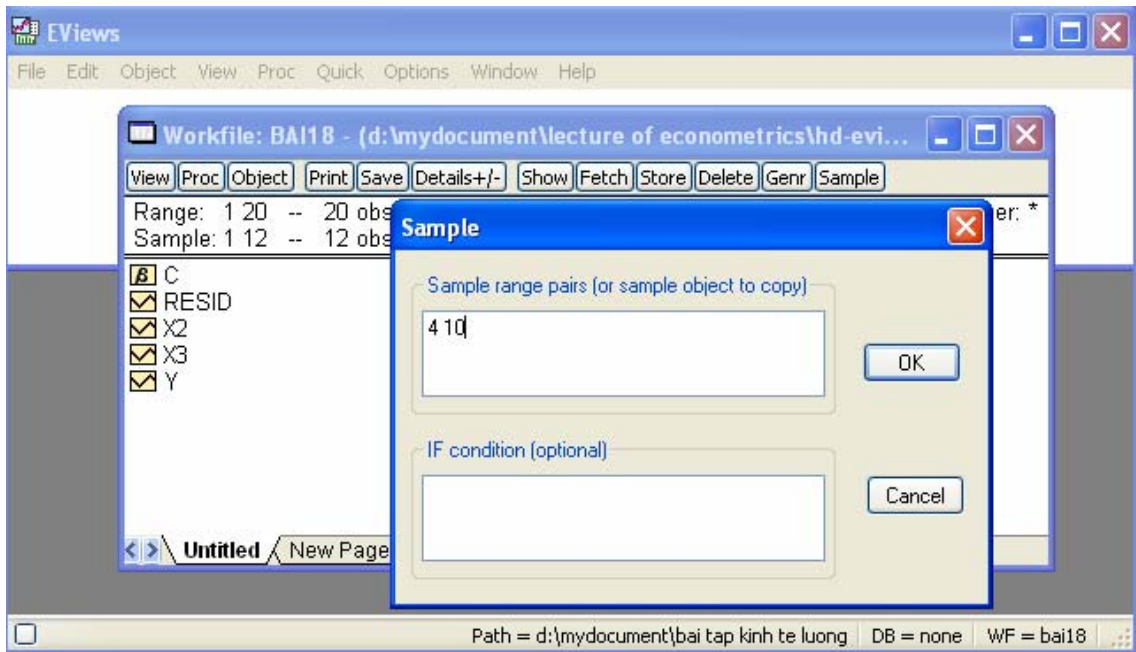


2.5. Lựa chọn mẫu dữ liệu (định mẫu mới).

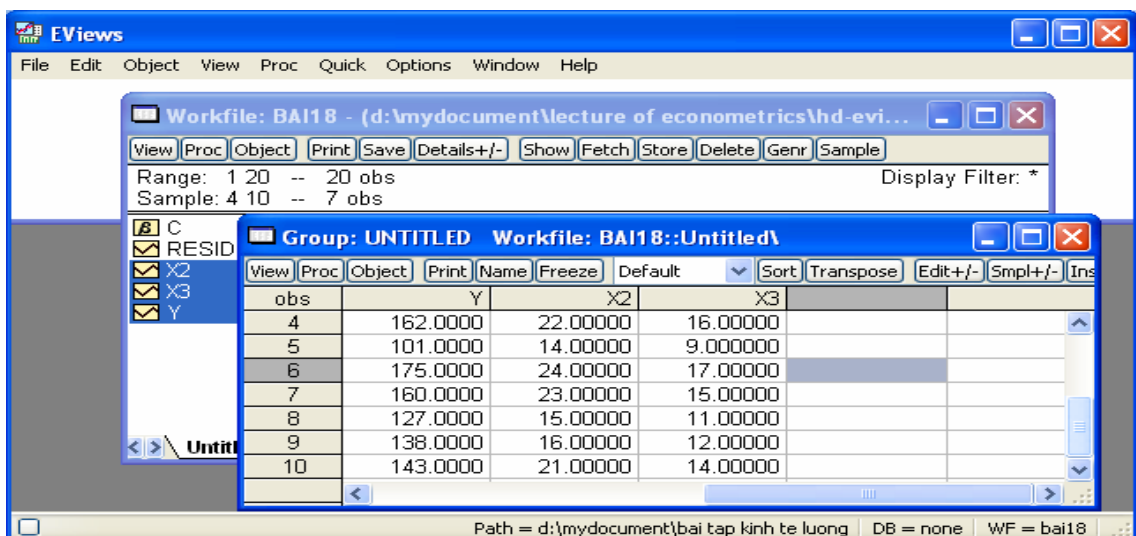
Lựa chọn mẫu dữ liệu là chọn giai đoạn nghiên cứu trong dữ liệu hiện có trong workfile. Chúng ta thường muốn tập trung phân tích vào một mẫu phụ của dữ liệu. EViews cho phép chúng ta làm điều này theo hai cách: (1) bằng cách xác định khoảng mẫu mà chúng ta muốn xem xét, (2) hoặc là bằng cách xác định các

quan sát thoả mãn một điều kiện logic (if) nào đó. Nếu chúng ta sử dụng cả hai phương pháp, thì mẫu kết quả là giao của các mẫu được tạo ra bởi hai phương pháp này.

Ví dụ, bây giờ ta muốn sử dụng một mẫu mới gồm 7 quan sát, từ quan sát thứ 4 đến quan sát thứ 10 trong mẫu này. Để làm việc này ta nhấp vào nút **Sample** trong bảng tính và đánh vào một mẫu mới: **4 10** (mỗi con số cách nhau ít nhất một khoảng trắng).



Sau khi đánh vào một mẫu mới theo mong muốn, bạn chọn **OK**, công việc sẽ hoàn tất. Để kiểm tra mẫu dữ liệu mới có đúng như mong muốn không, bạn hãy mở các biến ra theo nhóm sẽ thấy màn hình có kết quả như sau:

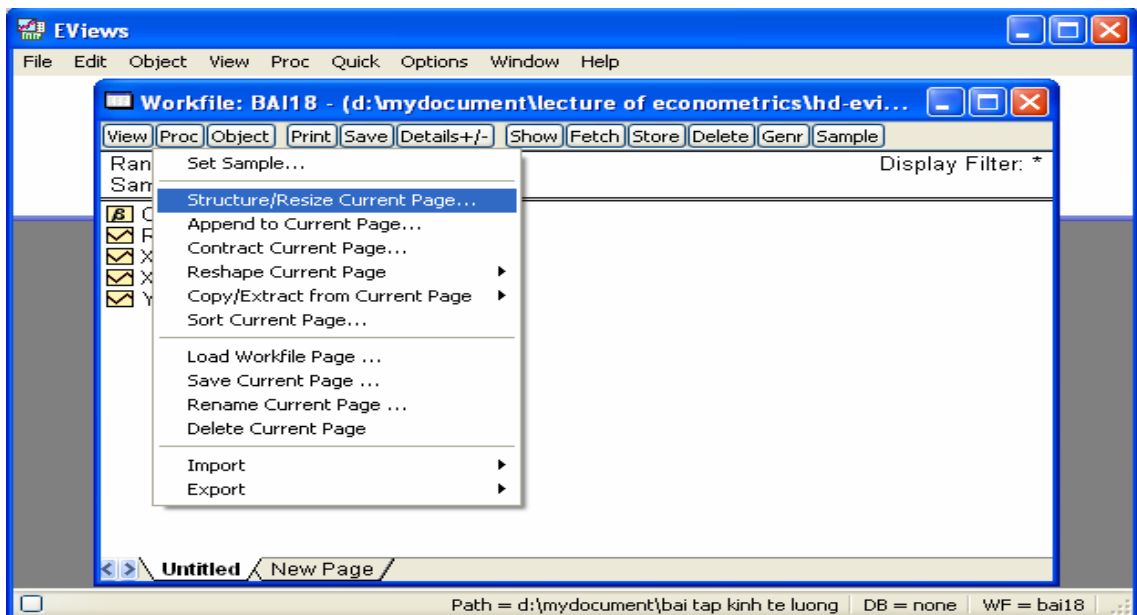


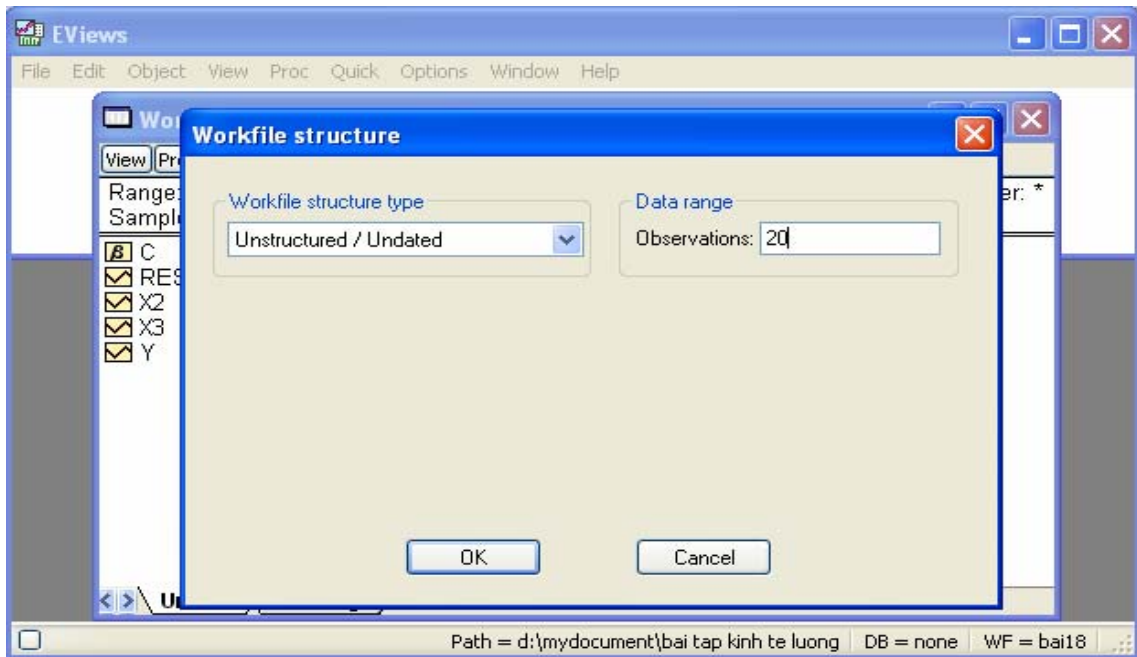
Nếu mẫu mới của bạn chỉ gồm những quan sát có Y lớn hơn 120 chẳng hạn, khi đó bạn gõ vào khung **IF condition (optional):** $Y > 120$. Hoặc nếu mẫu của bạn chỉ gồm những quan sát thỏa X_2 bằng 18, thì khi đó bạn đánh máy vào khung **IF condition (optional):** $X_2 = 18$.

2.6. Mở rộng Workfile làm việc.

Hãy tưởng tượng rằng chúng ta mong muốn nhận được dữ liệu bổ sung của 8 quan sát nữa cho workfile này. Chúng ta có thể mở rộng workfile này với mục đích bổ sung thêm dữ liệu này.

Nhấp **Proc/Structure Resize Current Page** và đánh máy vào khung **Data range** chỗ “observations” con số **20** (Vì workfile cũ hiện có 12 quan sát, ta muốn bổ sung thêm 8 quan sát nữa nên có tổng cộng là 20 quan sát). Bạn đọc cần lưu ý, nếu dữ liệu nghiên cứu là dữ liệu theo thời gian thì bạn chỉ cần đánh máy một **End Date (dữ liệu cuối)** mới.





Sau khi bạn chọn **OK** thì công việc đã hoàn tất. Tiếp tục mở nhóm gồm 3 biến Y, X2, và X3 ta sẽ có kết quả như sau:

obs	Y	X2	X3
9	138.0000	16.00000	12.00000
10	143.0000	21.00000	14.00000
11	158.0000	22.00000	15.00000
12	137.0000	13.00000	13.00000
13	NA	NA	NA
14	NA	NA	NA
15	NA	NA	NA
16	NA	NA	NA
17	NA	NA	NA
18	NA	NA	NA
19	NA	NA	NA
20	NA	NA	NA

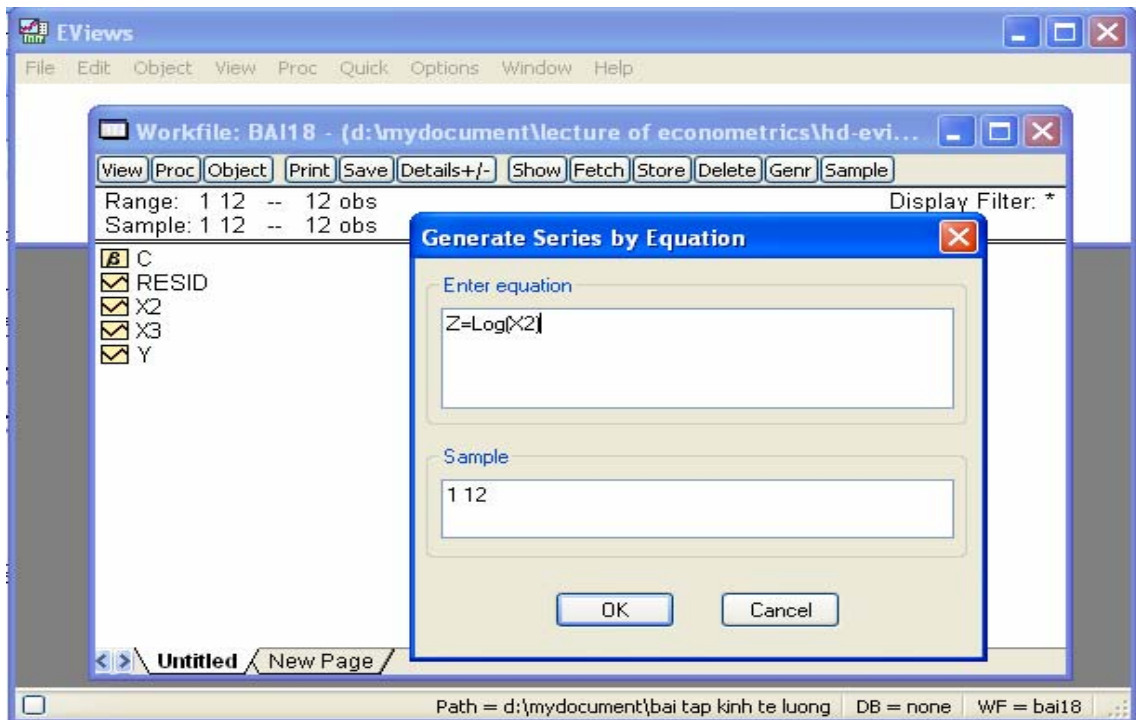
Nhìn vào bảng tính ta thấy từ quan sát thứ 13 đến 20 đều không có dữ liệu. Bây giờ bạn có thể biên tập bảng tính này để bổ sung thêm dữ liệu mới bằng cách nhập bằng tay từ bàn phím hoặc copy và dán thêm dữ liệu mới từ một tập tin Excel đã có sẵn chẳng hạn. Nên nhớ là khi nhập dữ liệu bổ sung bạn cần phải mở chức năng **Edit** trong Eviews nhé.

2.7. Tạo các biến mới, các độ trễ, các sai phân, xu hướng và các biến giả.

2.7.1. Tạo các biến mới.

Nếu chúng ta muốn tạo ra biến mới là một hàm số của các biến hiện có, thì chúng ta sử dụng chức năng **Genr** trên **Eviews**. Ví dụ, để tạo ra **Z** như một logarit tự nhiên của **X2** chúng ta sẽ nhấp vào **Genr** trên menu của workfile, sau đó đánh máy **Z = log(X2)**. Bạn cũng có thể tạo ra bất kỳ biến mới nào theo yêu cầu của bạn. Menu **Genr** có chức năng giống như một máy tính con, và còn hơn thế nữa.

Lưu ý rằng không bắt buộc phải tạo ra các biến mà bạn định sử dụng trong một phép hồi qui (hoặc qui trình khác), biểu thức có thể được đưa trực tiếp vào phần xác định phương trình hồi qui. Đặc tính này giúp ta giữ lại một workfile nhỏ bé.



Chọn **OK** là bạn đã hoàn tất công việc của mình. Máy tính sẽ tạo ra biến **Z** ngay trên Workfile của bạn. Bạn mở chuỗi **Z** ra và chọn **View/Descriptive Statistics/Histogram and Stats.**, khi đó màn hình sẽ như sau:

The screenshot shows the EViews interface with a workfile named 'BAI18'. A series 'Z' is selected, and its data is displayed in a spreadsheet view. The series is defined as $z = \log(x2)$. The data points are as follows:

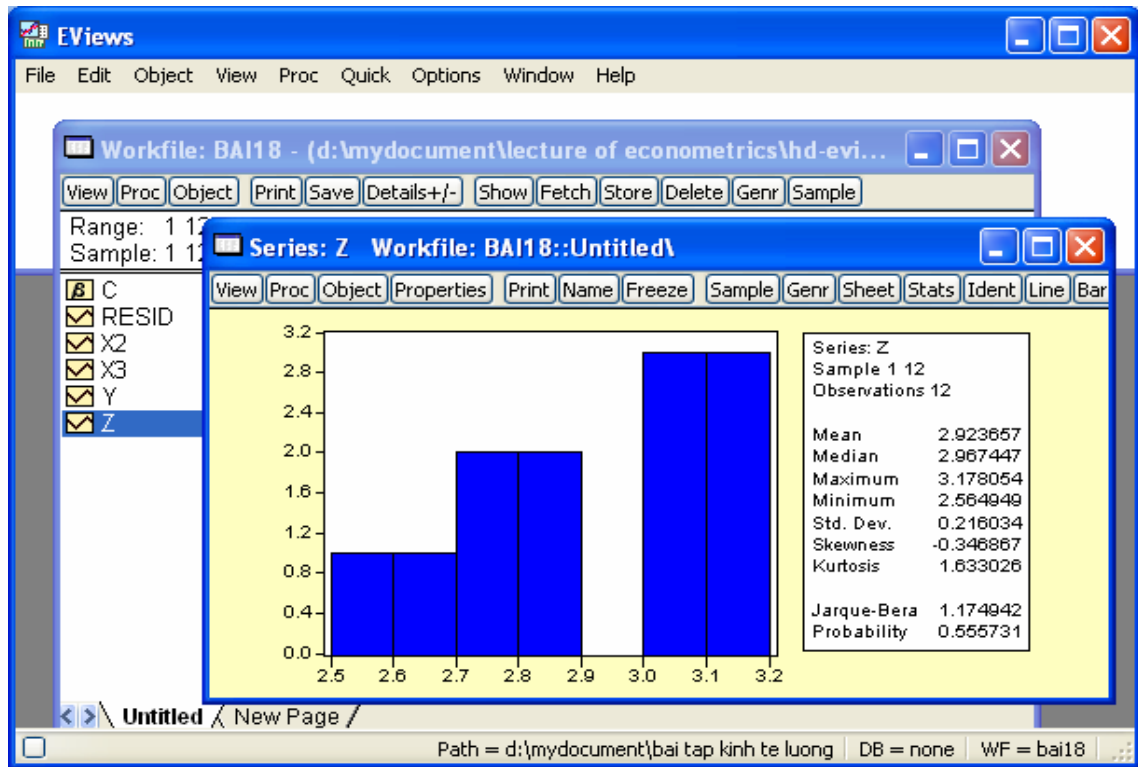
Series	Value
Z	2.833213
Z	3.135494
Z	2.890372
Z	3.091042
Z	2.639057
Z	3.178054
Z	3.135494

The status bar at the bottom shows the current value of the selected series: 2.89037175789616. The path is `d:\mydocument\bai tap kinh te luong`, the database is `DB = none`, and the workfile is `WF = bai18`.

The screenshot shows the EViews interface with the 'Descriptive Statistics' menu open for series 'Z'. The 'Histogram and Stats' option is highlighted. The menu options are:

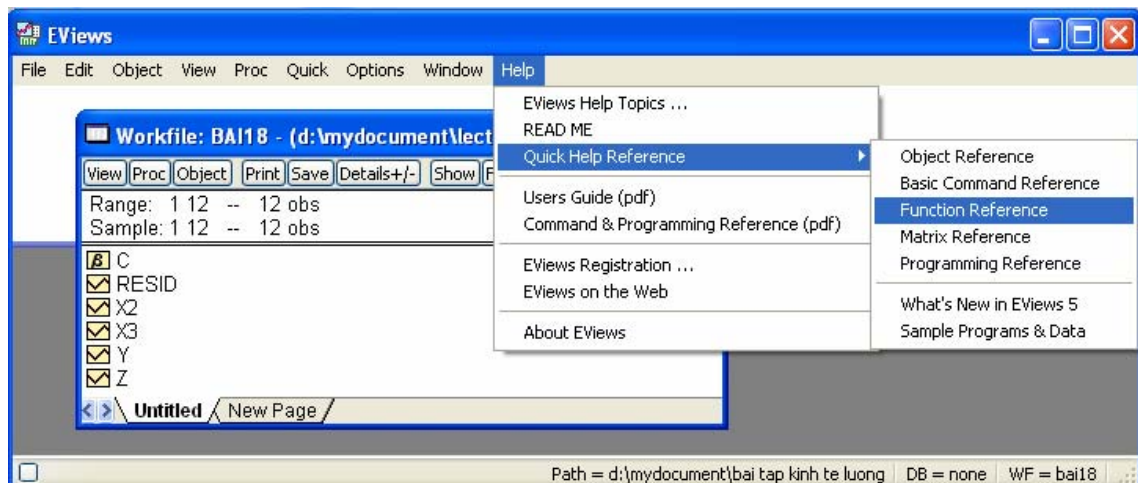
- SpreadSheet
- Graph
- Descriptive Statistics
 - Histogram and Stats
 - Stats Table
 - Stats by Classification...
 - Boxplots by Classification...
- Tests for Descriptive Stats
- Distribution
- One-Way Tabulation...
- Correlogram...
- Unit Root Test...
- BDS Independence Test...
- Properties...
- Label

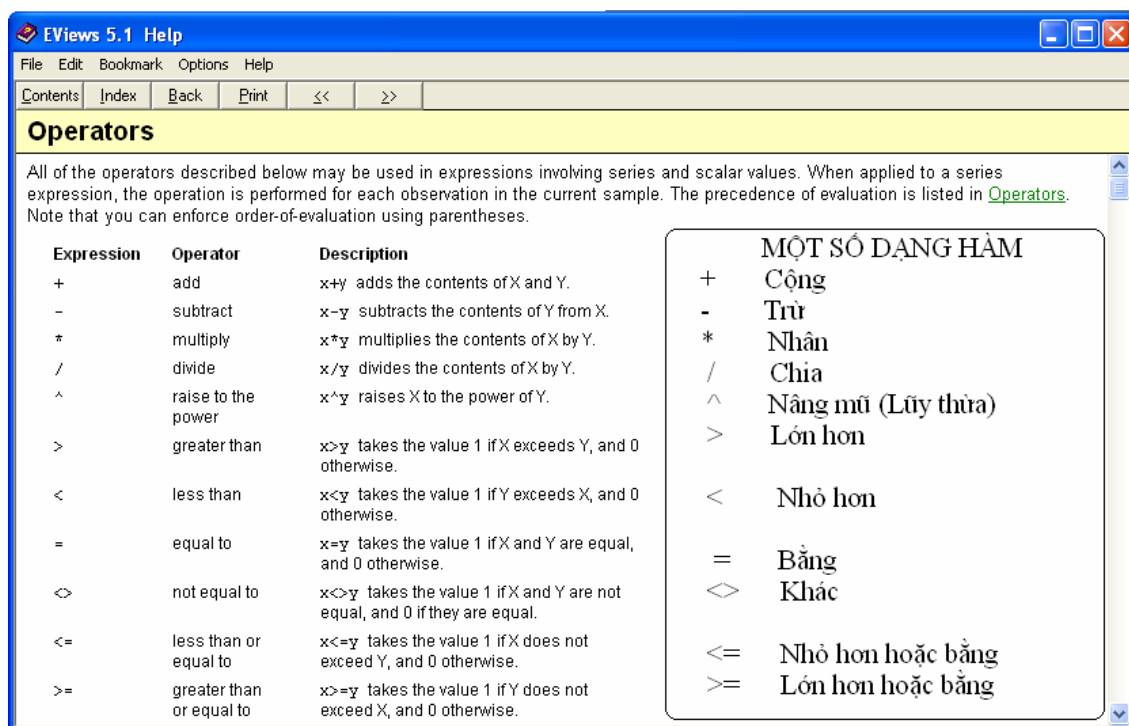
The status bar at the bottom shows the current value of the selected series: 2.89037175789616. The path is `d:\mydocument\bai tap kinh te luong`, the database is `DB = none`, and the workfile is `WF = bai18`.



Điều này cho thấy chúng ta có thể kiểm tra phân bố xác suất của một biến đơn như thế nào. Kết quả trên cho ta thấy biến **Z** có phân phối chuẩn, vì thống kê JB = 1.1749, với xác suất (P-value) = 0.5557 rất lớn nên ta không bác bỏ giả thuyết phân phối chuẩn của Z. Cửa sổ này sẽ rất hữu ích khi chúng ta xem xét các phần dư hồi qui.

Để tạo sự thuận lợi cho các bạn khi thực hiện các phép biến đổi, sau đây là một số hàm thông dụng trong Eviews mà ta thường sử dụng. EViews có một số lớn các chức năng mà bạn có thể khám phá dưới địa chỉ **Help/Function Reference**.





2.7.2. Tạo biến độ trễ và sai phân.

Trong nhiều ứng dụng chuỗi thời gian chúng ta đã sử dụng các biến có độ trễ, các sai phân bậc 1 (và bậc cao hơn), các biến giả có tính mùa vụ (hoặc theo tháng, quý,...), và các xu hướng thời gian. EViews có các lệnh rất dễ để tạo ra chuỗi theo ý muốn.

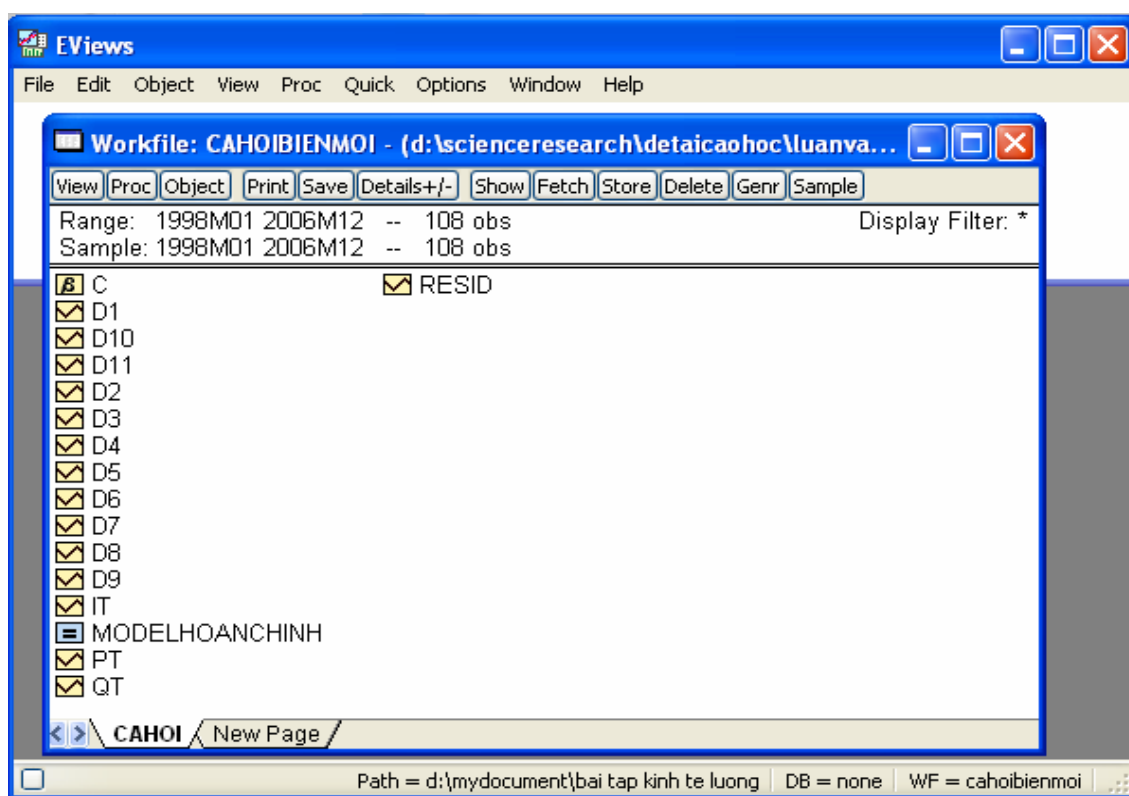
Để giúp bạn đọc có thể thực hiện một số công việc trong thực hành một cách dễ dàng, sau đây là một số hàm cơ bản trong Eviews.

Một số hàm trong Eviews	Ý nghĩa
X(-1)	Độ trễ của biến X (tạo ra chuỗi dữ liệu của X_{t-1})
D(X)	Sai phân bậc 1 của X ($D(X)=X_t-X_{t-1}$)
D(X,n)	Sai phân bậc n của X
LOG(X)	$\ln(X)$
EXP(X)	e^X (Hàm e^X)
ABS(X)	Giá trị tuyệt đối của X
SQR(X)	Căn bậc hai của X
@SUM(X)	Tổng của X
@SUMSQ(X)	Tổng của X^2

@MEAN(X)	Trung bình của X
@VAR(X)	Phương sai của X
@COV(X,Y)	Hiệp phương sai giữa X và Y
@COR((X,Y)	Hệ số tương quan giữa X và Y
@TREND(d)	Biến xu thế theo thời gian được chuẩn hóa về 0 ở thời kỳ d .
@SEAS(d)	Biến giả theo mùa bằng 1 khi quý hoặc tháng bằng d , bằng 0 nếu khác d .
@ABS(@QTDIST(α, df))	Giá trị tới hạn của đại lượng ngẫu nhiên t (phân phối student) với mức ý nghĩa α và bậc tự do df.
@TDIST(t_j , df)	Cho ta P-value ứng với 2 phía của thống kê t.
@QCHISQ(α, df)	Cho ta giá trị tới hạn của chi-bình phương với mức ý nghĩa α và bậc tự do df. (lưu ý: trong Eviews các giá trị tới hạn được tính theo đuôi bên trái, ngược lại với Excel).
@CHISQ(χ_0^2,df)	Cho ta P-value của chi-bình phương theo công thức $P\text{-value} = P(\chi^2 > \chi_0^2)$.
@QFDIST(α, df₁,df₂)	Cho ta giá trị tới hạn của đại lượng ngẫu nhiên Fisher với mức ý nghĩa α và bậc tự do của tử số là df ₁ , bậc tự do của mẫu là df ₂ . (lưu ý: trong Eviews các giá trị tới hạn được tính theo đuôi bên trái, ngược lại với Excel).
@FDIST(F_C, df₁,df₂)	Cho ta P-value của phân phối FISHER theo công thức $P\text{-value} = P(F > F_c)$.
MATRIX(5x5) X	Tạo ma trận cỡ 5x5 lấy tên là X
MATRIX C = a+b	Cộng 2 ma trận (hai ma trận cùng kích thước) matrix(3,4) a matrix(3,4) b
MATRIX C = a-b	Trừ 2 ma trận (hai ma trận cùng kích thước) matrix(3,4) a matrix(3,4) b
MATRIX C = a*b	Nhân 2 ma trận (ma trận a cấp nxm và b cấp

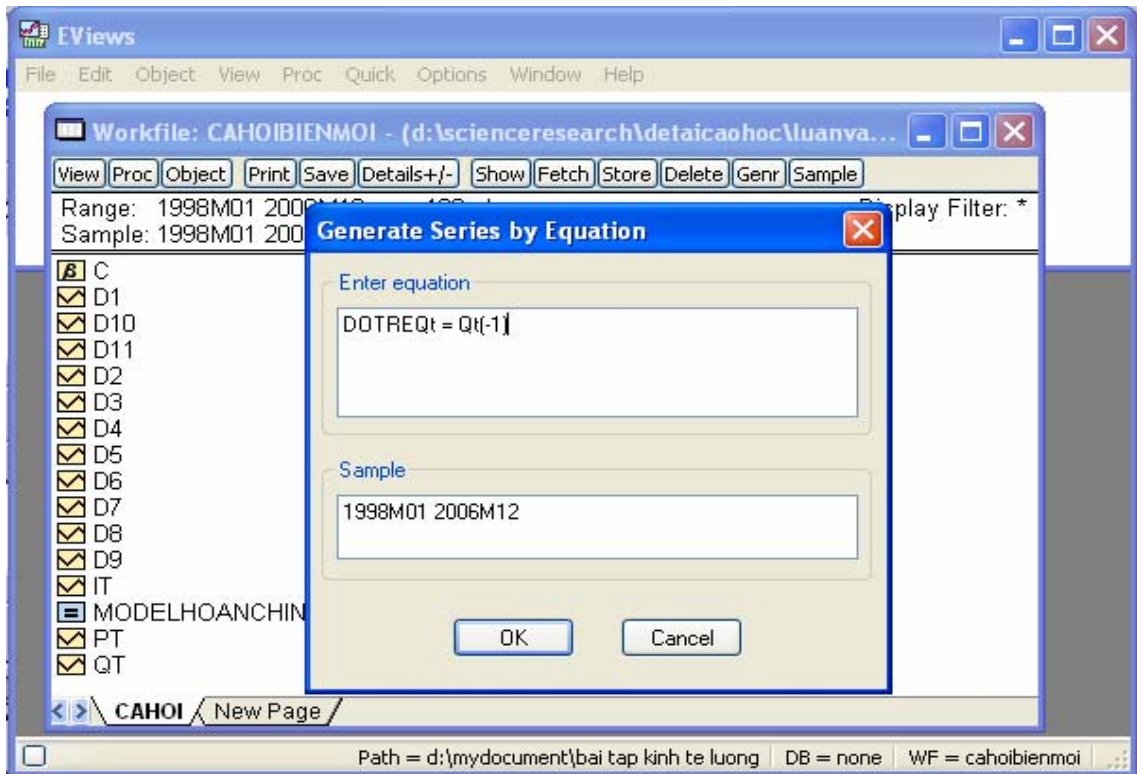
	mxq) matrix(5,9) a matrix(9,10) b
MATRIX W = @TRANPOSE(Y)	Chuyển vị của ma trận Y
MATRIX W = @INVERSE(Y)	Nghịch đảo của ma trận Y
MATRIX C = 3.14159*A MATRIX Z = A/3	Nhân, chia một số (scalar) vào ma trận tên A
@DET(A)	Định thức của ma trận A
Scalar X (scalar_name) = <Biểu thức tính toán hay Giá trị>	Scalar (single number). Chứa một giá trị nào đó do ta nhập vào hay là kết quả tính toán từ một biểu thức nào đó. Vd: Scalar pi = 3.14159 Scalar x = 3*y + 4

Workfile **CAHOIBIENMOI.wf1** có chứa dữ liệu về nhu cầu cá hồi (**Qt**), giá của cá hồi (**Pt**), thu nhập bình quân đầu người (**It**),... ở Việt Nam.

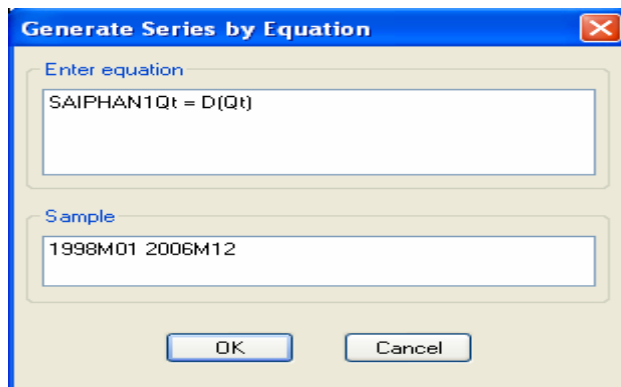


Quan sát đầu tiên là 1998M01 (tháng 01 năm 1998) và quan sát cuối cùng là 2006M12 (tháng 12 năm 2006).

Giá trị có độ trễ của **nhu cầu cá hồi (Qt)** có thể được tạo ra bằng cách nhập **Genr**, sau đó đánh máy vào phương trình: **DOTREQt = Qt(-1)**.



Để tạo ra sai phân thứ nhất của **cầu cá hồi (Qt)**, hãy nhấp **Genr**, và đánh máy vào phương trình: **SAIPHAN1Qt = D(Qt)**.



Bôi đen **Qt**, **DOTREQt** và **SAIPHAN1Qt** và kiểm tra dạng bảng tính của chúng, rất dễ thấy các mối tương quan giữa chúng.

The screenshot shows the EViews software interface. The main window displays a workfile named 'CAHOIBIENMOI' with a sample range from 1998M01 to 1999M02. The 'Group: UNTITLED' window is open, showing a list of variables and their values. The variables are: obs, QT, DOTREQT, SAIPHAN1QT, and SAIPHAN2QT. The values for each variable are listed in the table below.

Variable	Value
obs	1998M01
QT	1.000000
DOTREQT	NA
SAIPHAN1QT	NA
SAIPHAN2QT	NA
obs	1998M02
QT	17.000000
DOTREQT	1.000000
SAIPHAN1QT	16.000000
SAIPHAN2QT	-15.000000
obs	1998M03
QT	2.000000
DOTREQT	17.000000
SAIPHAN1QT	9.000000
SAIPHAN2QT	-10.000000
obs	1998M04
QT	11.000000
DOTREQT	2.000000
SAIPHAN1QT	1.000000
SAIPHAN2QT	0.000000
obs	1998M05
QT	1.000000
DOTREQT	11.000000
SAIPHAN1QT	1.000000
SAIPHAN2QT	0.000000
obs	1998M06
QT	2.000000
DOTREQT	1.000000
SAIPHAN1QT	0.000000
SAIPHAN2QT	0.000000
obs	1998M07
QT	2.000000
DOTREQT	2.000000
SAIPHAN1QT	0.000000
SAIPHAN2QT	0.000000
obs	1998M08
QT	1.000000
DOTREQT	2.000000
SAIPHAN1QT	-1.000000
SAIPHAN2QT	0.000000
obs	1998M09
QT	2.000000
DOTREQT	1.000000
SAIPHAN1QT	1.000000
SAIPHAN2QT	0.000000
obs	1998M10
QT	2.000000
DOTREQT	2.000000
SAIPHAN1QT	0.000000
SAIPHAN2QT	0.000000
obs	1998M11
QT	2.000000
DOTREQT	2.000000
SAIPHAN1QT	0.000000
SAIPHAN2QT	0.000000
obs	1998M12
QT	1.000000
DOTREQT	2.000000
SAIPHAN1QT	-1.000000
SAIPHAN2QT	0.000000
obs	1999M01
QT	2.000000
DOTREQT	1.000000
SAIPHAN1QT	1.000000
SAIPHAN2QT	0.000000
obs	1999M02
QT	2.000000
DOTREQT	1.000000
SAIPHAN1QT	1.000000
SAIPHAN2QT	0.000000

Các độ trễ có độ dài j có thể được tạo ra bằng cách sử dụng $Qt(-j)$ và các sai phân bậc j có thể được tạo ra bằng cách sử dụng: $SAIPHANjQt = Qt - Qt(-j)$.

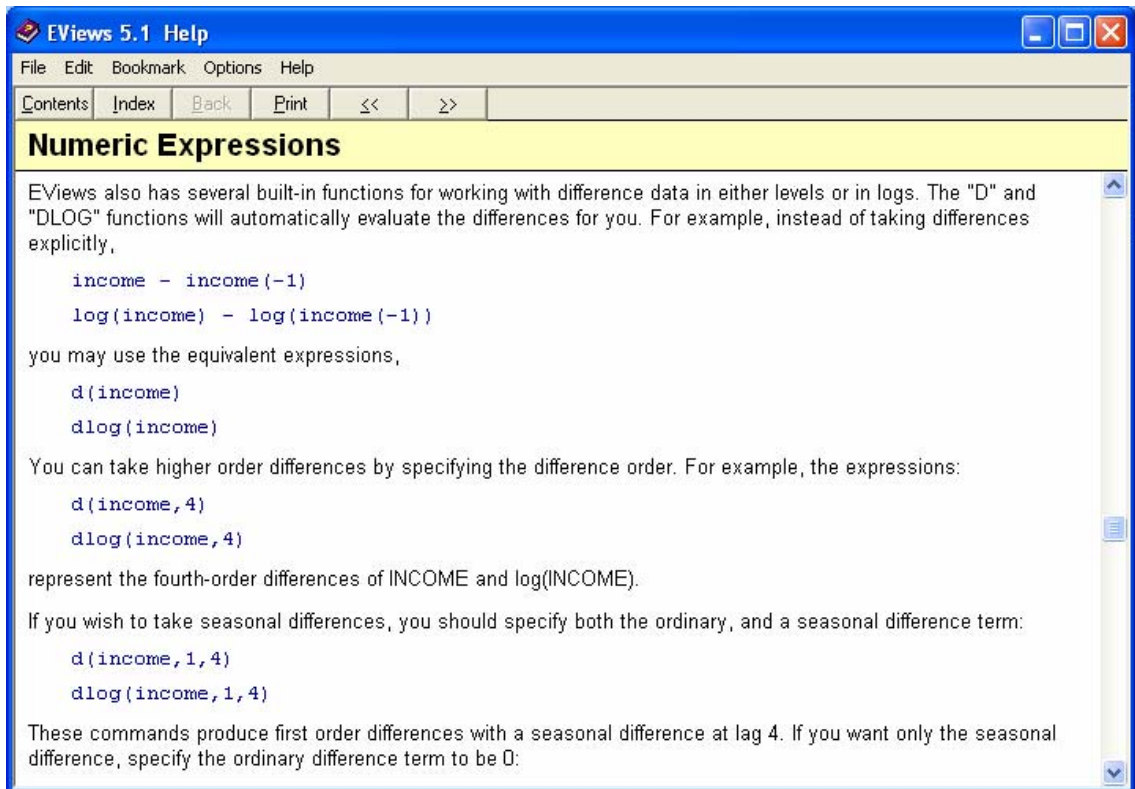
Các sai phân bậc cao hơn được tạo ra bởi toán tử sai phân $D(Qt, j)$. Ví dụ nếu $j = 2$, thì toán tử này sẽ tạo ra sai phân của các sai phân bậc nhất.

$$D(Qt, 2) = [Qt - Qt(-1)] - [Qt(-1) - Qt(-2)]$$

☛ **Công thức tạo biến mới chi tiết trong Eviews:**

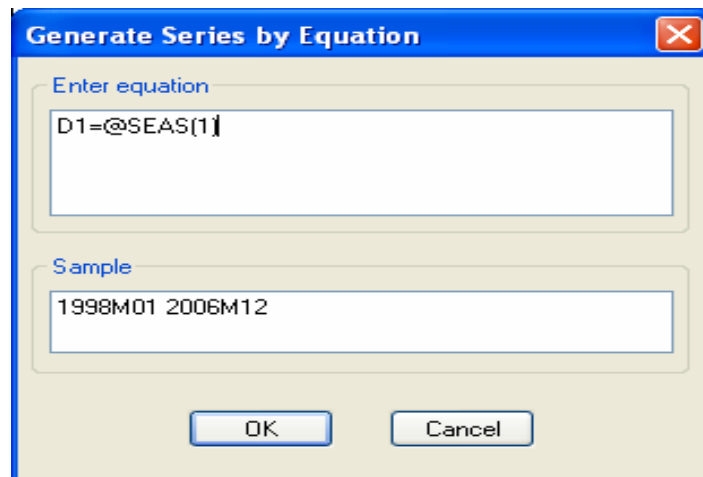
Nhằm đi xa hơn nữa các công thức tạo biến phức tạp, nếu không có hướng dẫn của giáo viên hoặc các chuyên gia phần mềm Eviews các bạn cần sử dụng lệnh sau đây: **Help / EViews Help Topics / Index**

Sau đó đánh vào Expressions. Khi chữ Expressions được bôi đen, nhấn **Display** và đọc các chỉ dẫn hiện ra.



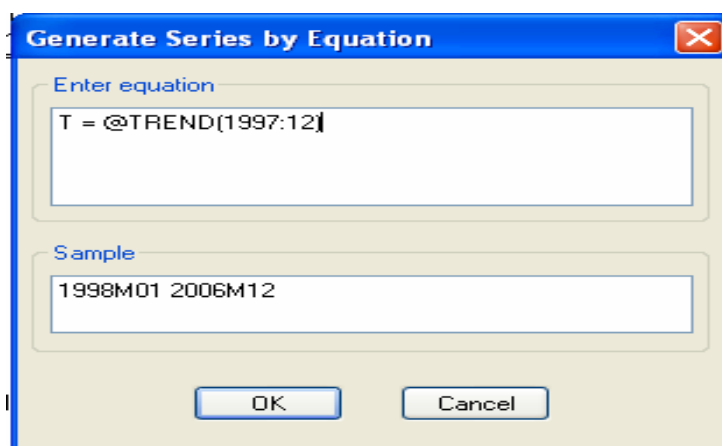
2.7.3. Tạo biến giả và biến xu hướng.

EViews có một số chức năng đặc biệt để tạo ra các biến giả có tính mùa vụ và các xu hướng thời gian. Nhấn **Genr** trên menu của workfile, sau đó đánh máy phương trình: **D1 = @SEAS(1)**



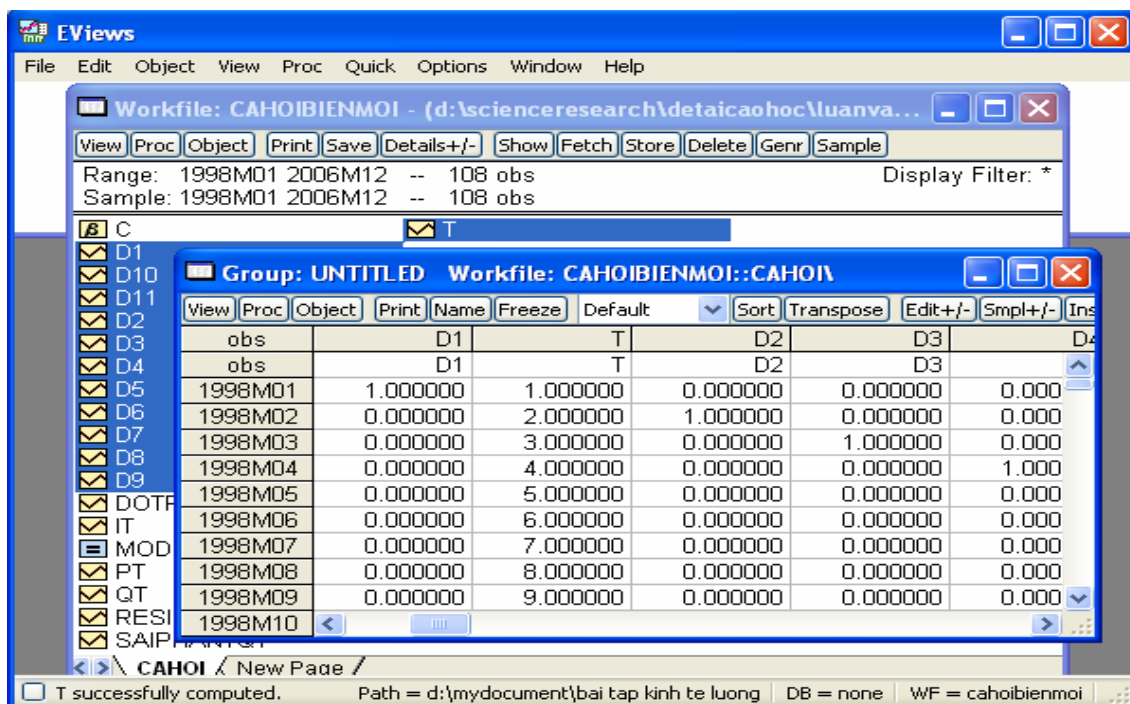
Việc này tạo ra một biến giả theo tháng (D1) có giá trị là 1 đối với tháng thứ nhất của mỗi năm và giá trị 0 đối với các tháng khác. Tạo **D2**, **D3**, **D4**,...bằng cách sử dụng cùng phương pháp y như vậy.

Để tạo ra biến xu hướng ta làm như sau: Nhấp **Genr** trên menu của workfile, sau đó đánh vào phương trình: **T = @TREND(1989:12)**



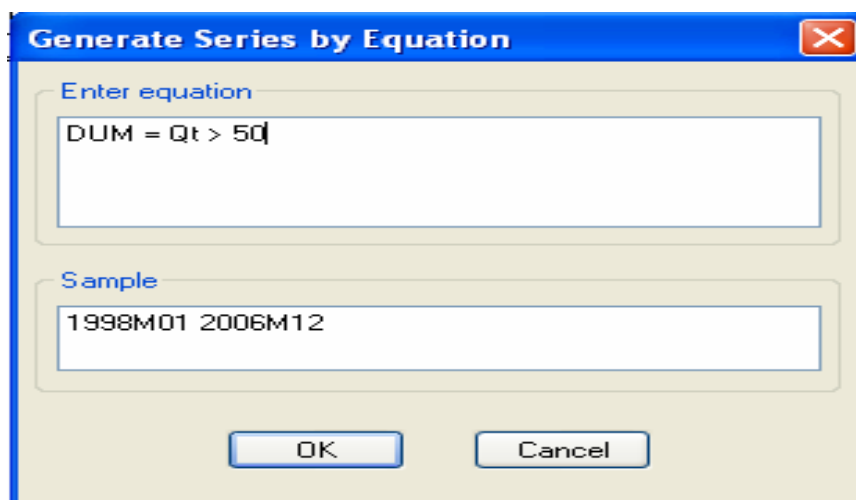
Việc này tạo ra một xu hướng thời gian có giá trị là 1 trong tháng đầu của năm 1998 và tăng thêm 1 sau mỗi tháng. Khi bạn sử dụng toán tử **@TREND(d)**, tham số **d** chỉ ra giai đoạn mà đối với nó giá trị của biến xu hướng bằng zero.

Để kiểm tra xem các biến giả theo tháng và biến xu hướng đã được tạo ra như thế nào, bạn hãy bôi đen **D1, D2, D3, D4, D5, D6,...** và **T**, rồi kiểm tra dạng bảng tính của nhóm này.



Ở trên đã ta nói việc tạo ra các biến giả có tính thời vụ hoặc theo thời gian. Tuy nhiên, Eviews cũng có thể tạo ra các biến giả **cho dữ liệu chéo** và các biến giả được sử dụng để đại diện cho sự có mặt hoặc là vắng mặt của các thuộc tính định lượng hoặc định tính.

Ví dụ, ta muốn tạo ra một biến giả có tên là **DUM** để xác định những quan sát (tháng) mà có nhu cầu (Qt) lớn hơn 50 chẳng hạn. Để làm việc này bạn chọn **Genr** trên menu của workfile và đánh vào phương trình: **DUM = Qt > 50**.



Việc này tạo ra một biến giả có giá trị bằng 1 khi nhu cầu cá hồi lớn hơn 50 ($Qt > 50$) và bằng 0 nếu không phải như vậy.

Các bạn có thể mở biến Qt và DUM theo nhóm để kiểm tra xem có đúng như vậy không nhé.

obs	QT	DUM
2004M07	44.00000	0.000000
2004M08	70.00000	1.000000
2004M09	30.00000	0.000000
2004M10	40.00000	0.000000
2004M11	55.00000	1.000000
2004M12	73.00000	1.000000
2005M01	120.0000	1.000000
2005M02	35.00000	0.000000
2005M03	95.00000	1.000000
2005M04	81.00000	1.000000
2005M05		

Bây giờ, ta lấy một ví dụ về dữ liệu chéo để minh họa cho việc tạo ra các biến giả.

Để nghiên cứu nhu cầu của một loại hàng hóa, người ta đã tiến hành khảo sát giá cả và lượng hàng tiêu thụ được ở 10 cửa hàng như sau:

Lượng hàng bán được (Y - tấn/tháng)	Giá bán (X - 1000đ/kg)	Khu vực bán
30,0	2,0	Nông thôn
29,5	2,1	Thành thị
29,2	2,2	Nông thôn
28,2	2,3	Thành thị
28,9	2,4	Thành thị
28,5	2,5	Nông thôn
28,2	2,6	Nông thôn
27,7	2,7	Thành thị
27,3	2,8	Thành thị
27,5	2,9	Nông thôn

Hãy xây dựng biến giả cho biến "khu vực", với thành thị là phạm trù điều kiện.

Dữ liệu của ví dụ trên được nhập vào Eviews, màn hình như sau:

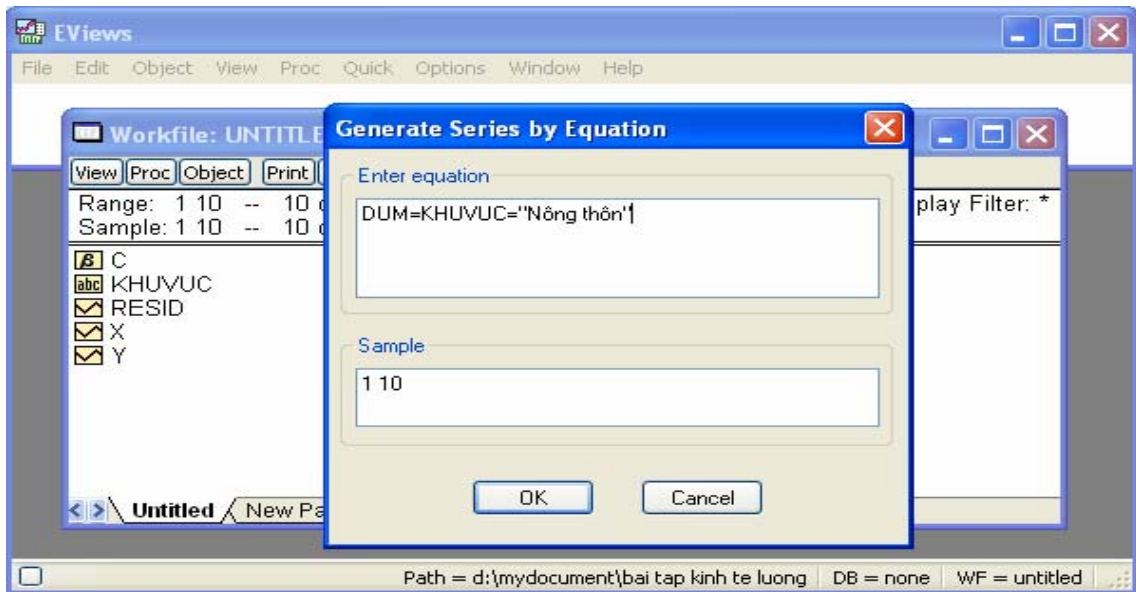
The screenshot shows the EViews interface. The main window displays a workfile with 10 observations. The 'Group: UNTITLED' window is open, showing a table with the following data:

obs	Y	X	KHUVUC
1	30.00000	2.000000	Nông thôn
2	29.50000	2.100000	Thành thị
3	29.20000	2.200000	Nông thôn
4	28.20000	2.300000	Thành thị
5	28.90000	2.400000	Thành thị
6	28.50000	2.500000	Nông thôn
7	28.20000	2.600000	Nông thôn
8	27.70000	2.700000	Thành thị
9	27.30000	2.800000	Thành thị
10			

Lưu ý: Biến KHUVUC không phải là một chuỗi số mà là một chuỗi vừa là ký số vừa là ký tự (Alphanumeric).

Để xây dựng biến giả cho biến "khu vực", với thành thị là phạm trù điều kiện, có nghĩa là "thành thị" sẽ nhận giá trị bằng 0.

Từ cửa sổ Workfile bạn chọn **Genr** rồi đánh vào câu lệnh sau:
DUM=KHUVUC= “Nông thôn”.



Việc này tạo ra một biến giả có tên là DUM với “**Nông thôn**” nhận giá trị bằng 1, và bằng 0 nếu không phải là “**Nông thôn**”. Bây giờ bạn mở hai biến KHUVUC và DUM ra thành một nhóm và kiểm tra chúng.

obs	KHUVUC	DUM
1	Nông thôn	1.000000
2	Thành thị	0.000000
3	Nông thôn	1.000000
4	Thành thị	0.000000
5	Thành thị	0.000000
6	Nông thôn	1.000000
7	Nông thôn	1.000000
8	Thành thị	0.000000
9	Thành thị	0.000000
10	Nông thôn	1.000000

2.8. Ước lượng mô hình hồi quy:

Để thuận lợi cho các bạn trong việc thực hiện phép ước lượng mô hình hồi quy, sau đây là một số dạng hàm hồi quy thông dụng và câu lệnh tương ứng dùng để ước lượng mô hình.

MỘT SỐ DẠNG HÀM HỒI QUY	CÂU LỆNH
$Y = \beta_1 + \beta_2 X_2 + \dots + \beta_K X_K + U$	LS Y C X ₂ ...X _K
$\ln(Y) = \beta_1 + \beta_2 \ln(K) + \beta_3 \ln(L) + U$	LS Log(Y) C Log(K) Log(L)
$\ln(Y) = \beta_1 + \beta_2 X_t + U$	LS Log(Y) C X
$Y = \beta_1 + \beta_2 \ln(X) + U$	LS Y C Log(X)
$Y = \beta_1 + \beta_2 1/X + U$	LS Y C 1/X
$Y = \beta_0 + \beta_1 X + \beta_2 X^2 + \dots + \beta_K X^K + U$	LS Y C X X ² ...X ^K
$\ln(e_i^2) = \beta_1 + \beta_2 \ln(X_i) + v_i$	LS Log(Resid ²) C Log(X)
$ e_i = \beta_1 + \beta_2 X_i + v_i$	LS ABS(Resid) C X
$ e_i = \beta_1 + \beta_2 \sqrt{X} + v_i$	LS ABS(Resid) C Sqr(X)
$ e_i = \beta_1 + \beta_2 1/X + v_i$	LS ABS(Resid) C 1/X
$ e_i = \beta_1 + \beta_2 1/\sqrt{X} + v_i$	LS ABS(Resid) C 1/Sqr(X)

Ghi chú: e_i là phần dư (trong evIEWS được ký hiệu là **Resid**).

LS: Viết tắt chữ Least Square (Phương pháp OLS).

C: Vector hệ số - tương ứng với hệ số tung độ gốc.

Như vậy, trong phần câu lệnh chúng ta thấy với mô hình hồi quy như thế nào thì câu lệnh được viết tương ứng. Đầu tiên là chúng ta khai báo phương pháp ước lượng bình phương bé nhất – **LS**, kế đến là biến phụ thuộc, vector hệ số (**C**), rồi đến các biến độc lập. Mỗi biến cách nhau ít nhất một khoảng trắng.

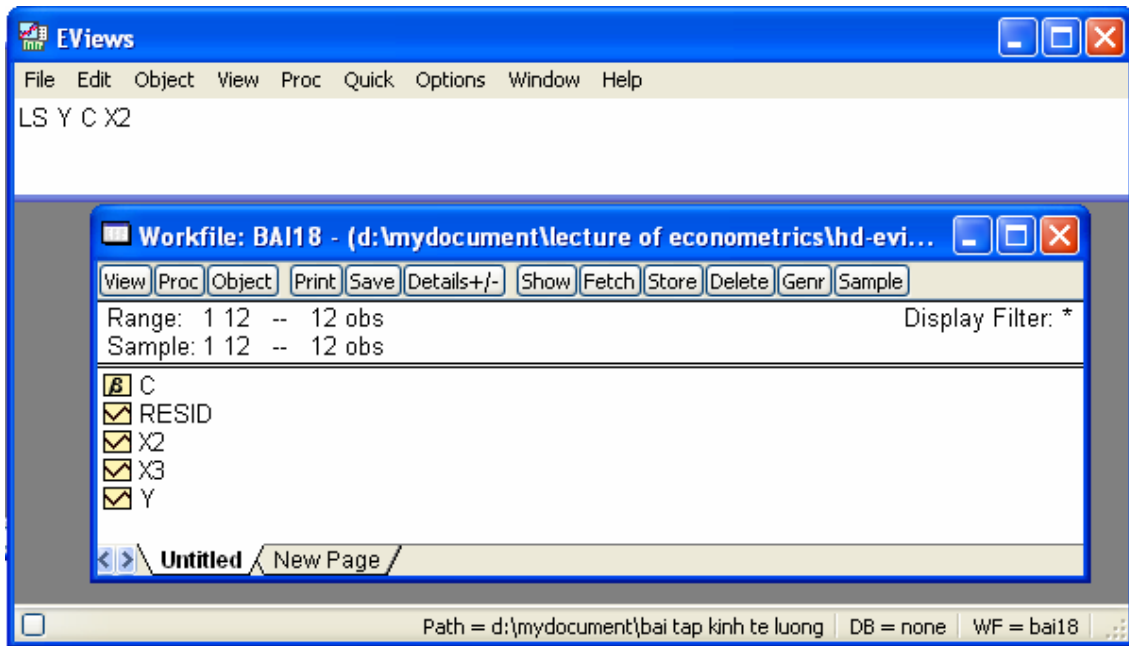
2.8.1. Ước lượng mô hình hồi quy đơn.

Bạn có thể ước lượng mô hình bằng nhiều cách khác nhau, sau đây là hai trong các cách để ước lượng mô hình hồi quy.

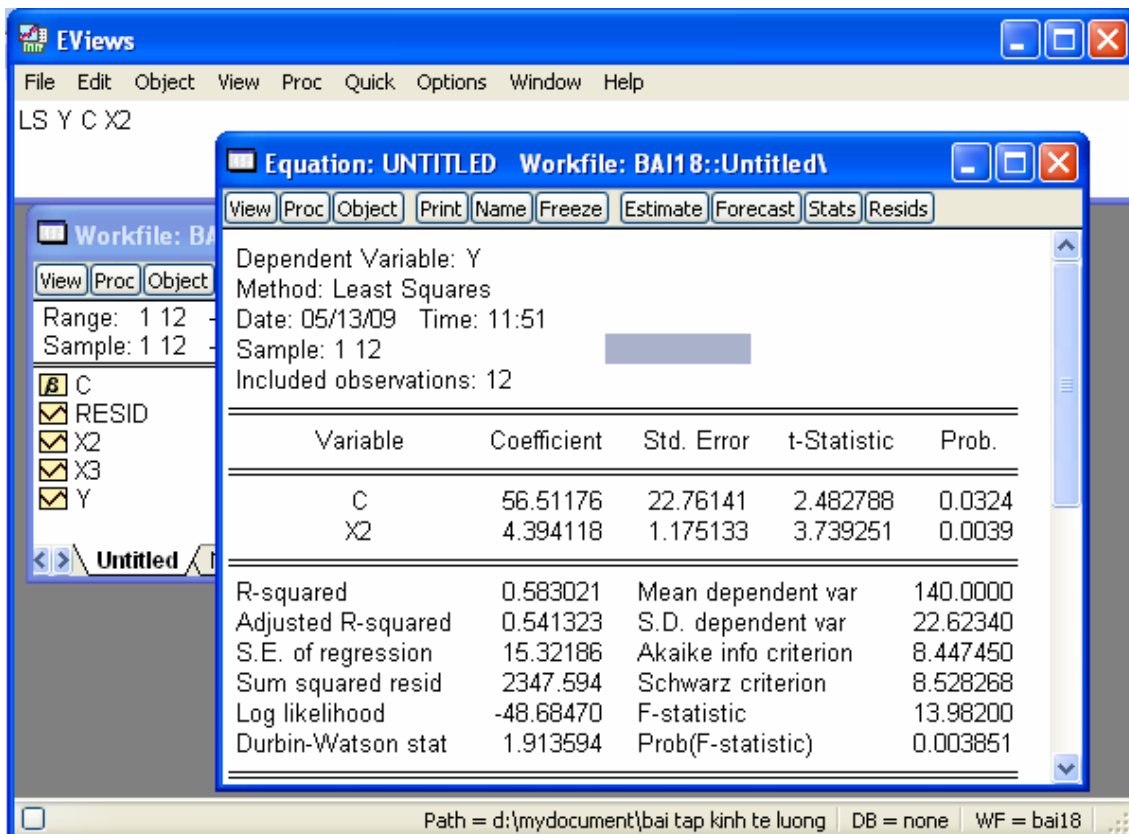
Hãy khởi động EVIEWS và mở workfile **Bai18.wf1** ra. Workfile này chứa dữ liệu của ba biến gồm: Y (Doanh số), X2 (Chi phí quảng cáo), X3 (Tiền lương của nhân viên tiếp thị) của 12 công ty.

Cách 1:

Giả sử ta muốn ước lượng mô hình hồi quy tuyến tính mẫu của Y (doanh số) theo X2 (chi phí quảng cáo). Mô hình có dạng như sau: SRF: $\hat{Y}_i = \hat{\beta}_1 + \hat{\beta}_2 X_{2i}$.

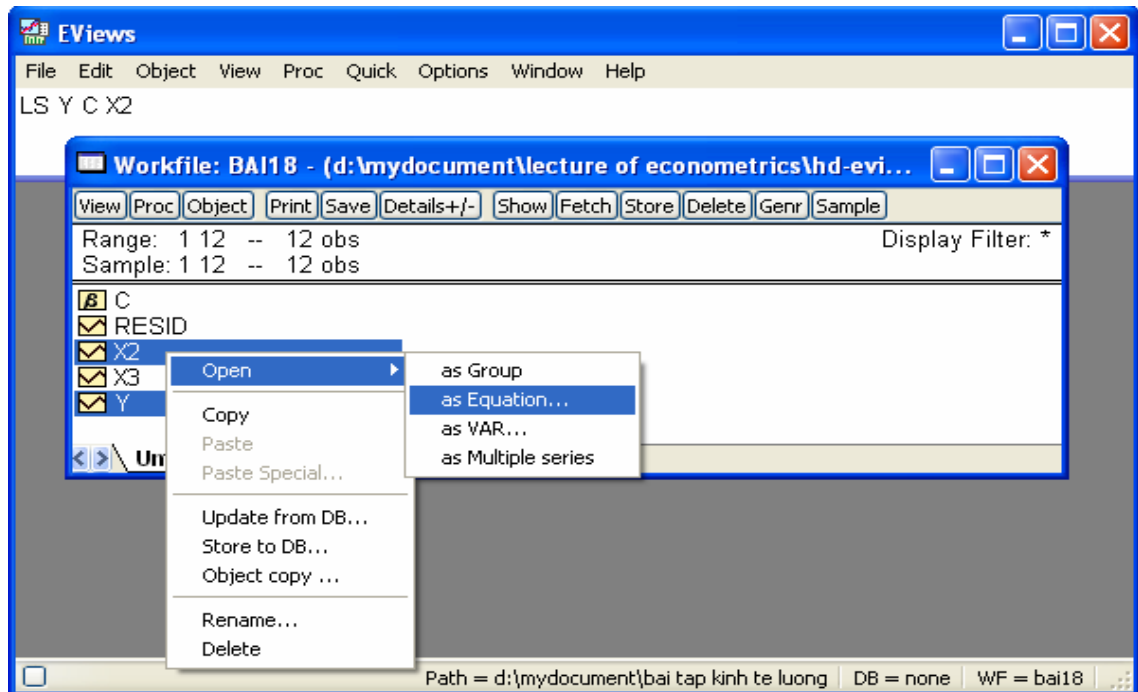


Trên cửa sổ lệnh của menu chính bạn đánh vào câu lệnh: **LS Y C X2** như ở hình trên rồi gõ **Enter**. Kết quả ước lượng sẽ như sau:

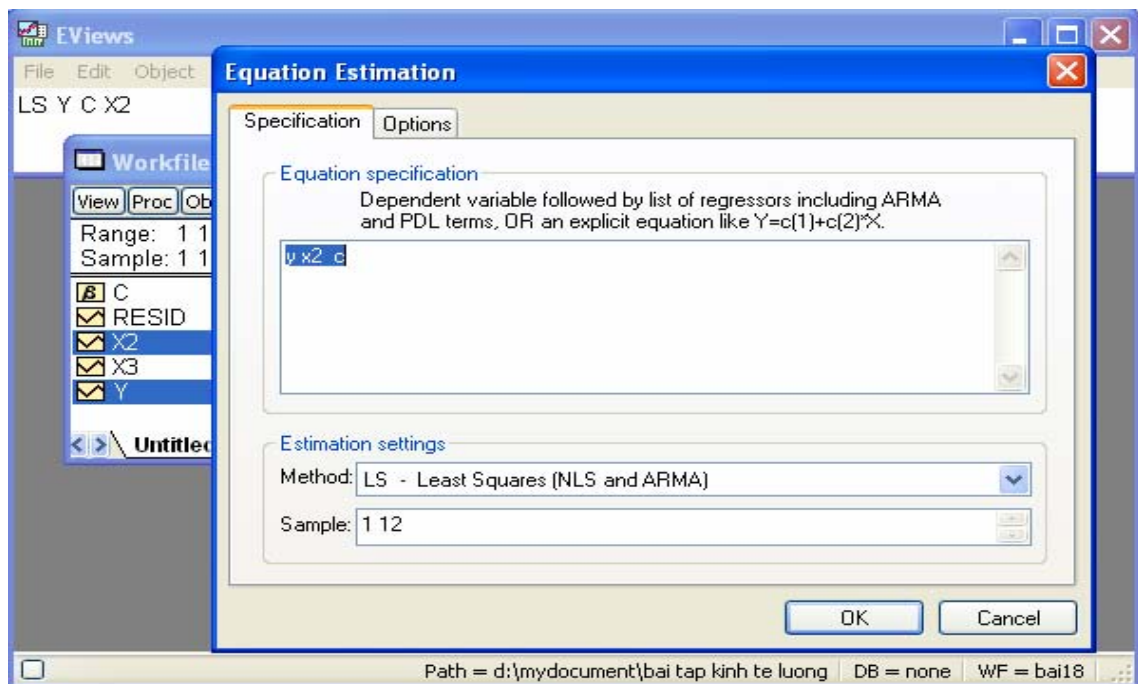


Cách 2:

Bạn bôi đen biến Y và X2(chọn biến phụ thuộc trước, rồi đến các biến độc lập), sau đó đưa chuột tới chỗ đã đánh dấu rồi bấm chuột phải, chọn **Open/ as Equation...** Bạn có thể nhìn thấy ở hình sau.



Khi bạn chọn **as Equation...** một màn hình sẽ xuất hiện như sau:

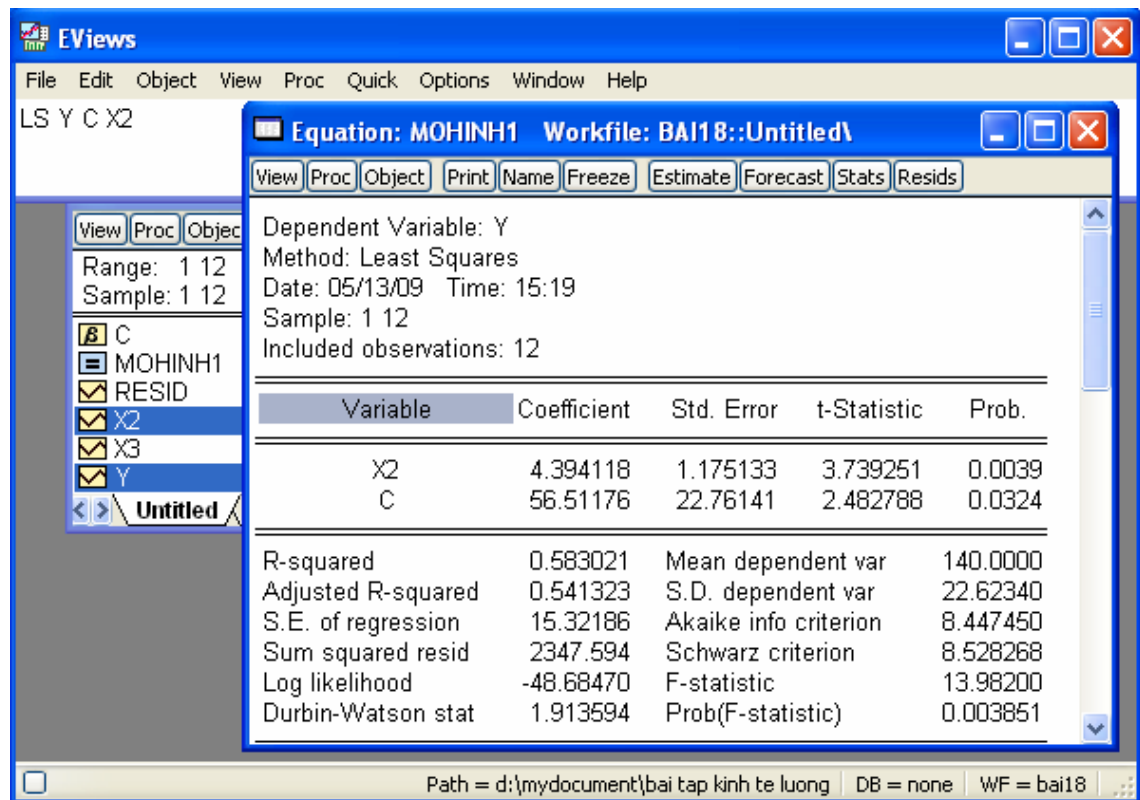


Trong khung **Equation specification**, bạn đánh câu lệnh vào. Ghi nhớ là biến đầu tiên bao giờ cũng là biến phụ thuộc, tiếp theo là các biến độc lập. Hệ số

tung độ gốc được Eviews ngầm định là C. Vector hệ số C có thể để ở sau cùng hoặc ngay sau biến phụ thuộc đều được.

Trong khung **Estimation Settings**, bạn cần phải khai báo vào chỗ “**Method**” phương pháp dùng để ước lượng. Ở đây ta sử dụng phương pháp OLS nên ta chọn “**LS – Least square...**”. Lựa chọn “**Sample**” cho ta biết mẫu ước lượng của ta là từ quan sát thứ 1 đến quan sát thứ 12. Nếu bạn muốn ước lượng mô hình nhưng không sử dụng hết các quan sát của mẫu thì bạn có thể điều chỉnh cho phù hợp với mẫu mà bạn muốn. Chẳng hạn, bạn muốn ước mô hình nhưng chỉ với các quan sát từ quan sát thứ 4 đến quan sát thứ 10 thì bạn thay đổi “**1 12**” thành “**4 10**”.

Sau khi khai báo xong, bạn chọn “**OK**”, kết quả ước lượng hiện ra trong bảng sau:



The screenshot shows the EViews software interface. The main window displays the equation **Equation: MOHINH1** and the workfile **Workfile: BAI18::Untitled1**. The regression results are as follows:

Variable	Coefficient	Std. Error	t-Statistic	Prob.
X2	4.394118	1.175133	3.739251	0.0039
C	56.51176	22.76141	2.482788	0.0324

Additional statistics shown in the window include:

- R-squared: 0.583021
- Adjusted R-squared: 0.541323
- S.E. of regression: 15.32186
- Sum squared resid: 2347.594
- Log likelihood: -48.68470
- Durbin-Watson stat: 1.913594
- Mean dependent var: 140.0000
- S.D. dependent var: 22.62340
- Akaike info criterion: 8.447450
- Schwarz criterion: 8.528268
- F-statistic: 13.98200
- Prob(F-statistic): 0.003851

Bạn có thể đặt tên cho mô hình này để phân biệt với các mô hình khác đã ước lượng hoặc sau này tiếp tục sử dụng đến nó. Muốn vậy, trên menu **Equation** bạn chọn **Name**, máy sẽ tự động tạo ra một tên khác với tên các phương trình đã có trước đó, theo thứ tự từ nhỏ đến lớn: EQ01, EQ02,... Dĩ nhiên, bạn có thể đặt một tên nào đó mà bạn thích. Ở đây tôi đặt tên là **MOHINH1**.

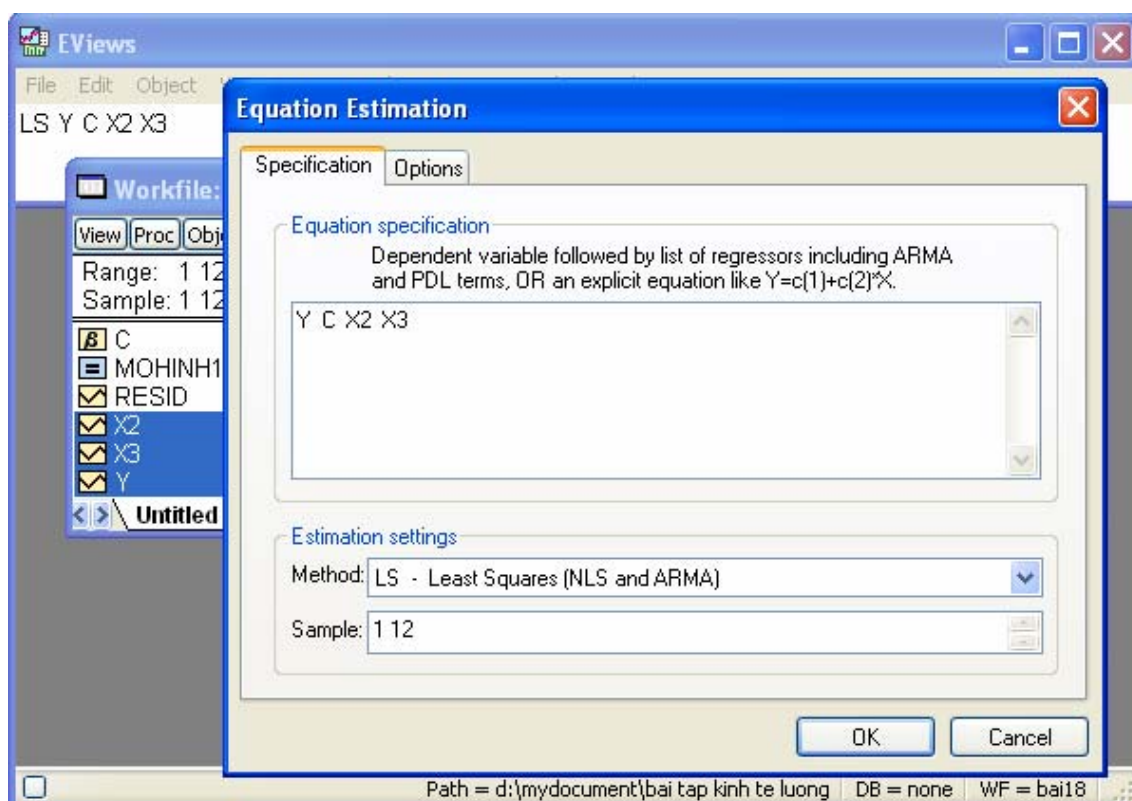
Cửa sổ **Equation: MOHINH1** cho ta kết quả tóm tắt của phân tích hồi quy. Các bạn giải thích kết quả ước lượng này và liên hệ nó với bài học của mình trên lớp nhé.

Dependent Variable: Y - Biến phụ thuộc Y				
Method: Least Squares – Phương pháp bình phương nhỏ nhất (OLS)				
Date: 05/13/09 Time: 15:19 – Ngày, giờ thực hiện hồi quy này.				
Sample: 1 12 – Mẫu ước lượng từ quan sát thứ 1 đến quan sát thứ 12.				
Included observations: 12 – Tổng số quan sát là 12				
Variable (Cột Biến độc lập)	Coefficient (Cột hệ số hồi quy)	Std. Error (Sai số chuẩn ước lượng của các hệ số hồi quy)	t-Statistic (Thống kê t)	Prob. (P-value của t)
X2	4.394118	1.175133	3.739251	0.0039
C	56.51176	22.76141	2.482788	0.0324
R-squared: 0.583021 (Hệ số xác định R ²)		Mean dependent var: 140.0000 (Trung bình của biến phụ thuộc)		
Adjusted R-squared: 0.541323 (R ² hiệu chỉnh)		S.D. dependent var: 22.62340 (Độ lệch chuẩn của biến phụ thuộc)		
S.E. of regression 15.32186 (Sai số chuẩn của hàm hồi quy mẫu)		Akaike info criterion: 8.447450 (Tiêu chuẩn Akaike – AIC)		
Sum squared resid: 2347.594 (Tổng bình phương các phần dư - RSS)		Schwarz criterion: 8.528268 (Tiêu chuẩn Schwarz – SC)		
Log likelihood: -48.68470 (Ln hàm hợp lý)		F-statistic: 13.98200 (Thống kê F)		
Durbin-Watson stat: 1.913594 (Thống kê Durbin-Watson)		Prob(F-statistic): 0.003851 (P-value của F)		

2.8.2. Ước lượng mô hình hồi quy bội.

Bây giờ, ta muốn ước lượng mô hình hồi quy tuyến tính mẫu của Y (doanh số) theo X2 (chi phí quảng cáo) và X3 (tiền lương của nhân viên tiếp thị). Khi đó mô hình hồi quy bội có dạng như sau: SRF: $\hat{Y}_i = \hat{\beta}_1 + \hat{\beta}_2 X_{2i} + \hat{\beta}_3 X_{3i}$.

Với mục đích ước lượng phương trình này, bạn có thể làm giống như khi ta ước lượng mô hình hồi quy đơn. Bạn chỉ việc gỡ thêm biến độc lập X3 vào hộp xác định phương trình (**Equation Specification**) như sau: **Y C X2 X3**.



Chọn “OK” kết quả như sau:

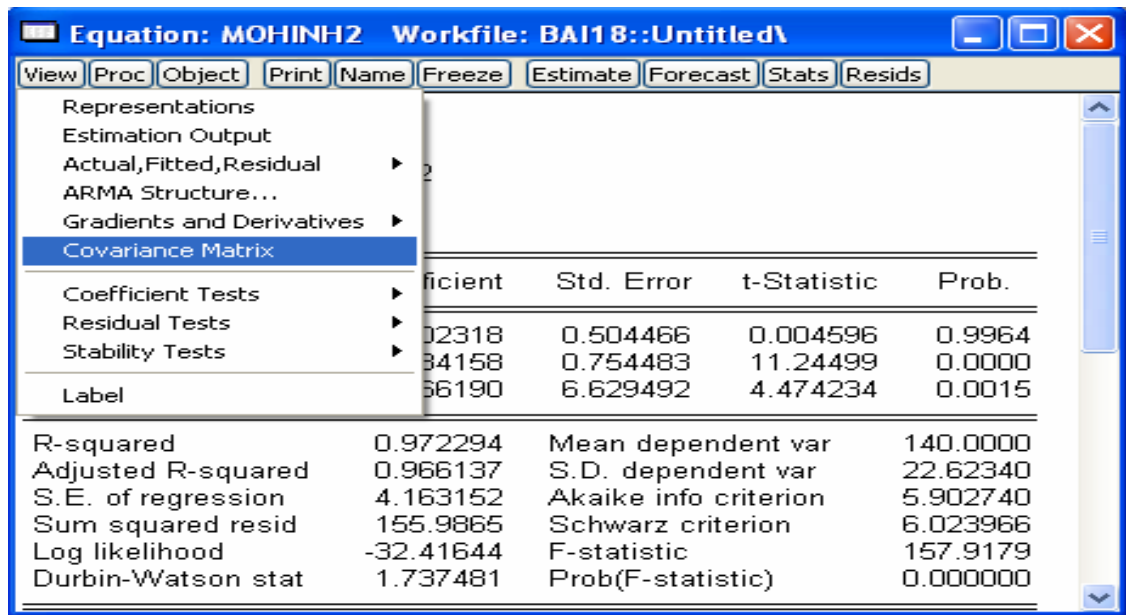
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	29.66190	6.629492	4.474234	0.0015
X2	0.002318	0.504466	0.004596	0.9964
X3	8.484158	0.754483	11.24499	0.0000
R-squared	0.972294	Mean dependent var		140.0000
Adjusted R-squared	0.966137	S.D. dependent var		22.62340
S.E. of regression	4.163152	Akaike info criterion		5.902740
Sum squared resid	155.9865	Schwarz criterion		6.023966
Log likelihood	-32.41644	F-statistic		157.9179
Durbin-Watson stat	1.737481	Prob(F-statistic)		0.000000

Các bạn có thể giải thích kết quả này tương tự như bảng kết quả trong mô hình hồi quy đơn.

2.9. Tìm ma trận hiệp phương sai của các hệ số hồi quy và một số giá trị khác như phần dư(e_i), giá trị Y_i ,...

Sau khi ước lượng mô hình hồi quy xong, chúng ta có thể xác định ma trận hiệp phương sai của các hệ số hồi quy như sau.

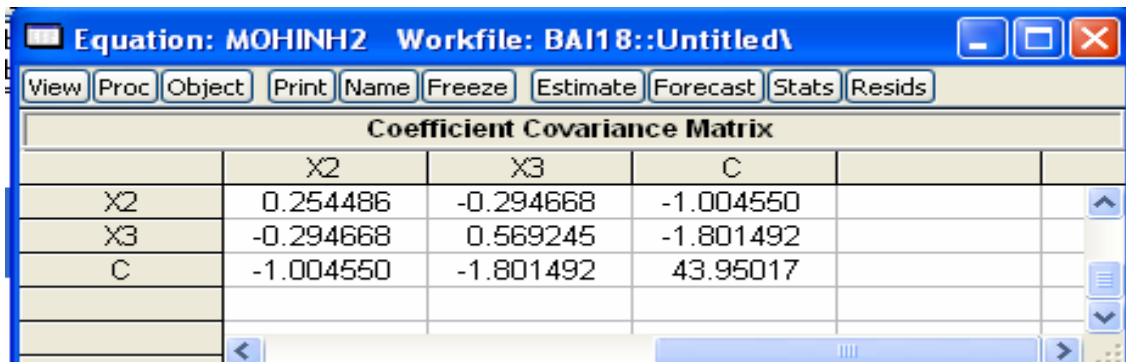
Từ cửa sổ Equation: MOHINH2 bạn chọn **View/Covariance Matrix**, khi đó màn hình như sau:



	Coefficient	Std. Error	t-Statistic	Prob.
	0.02318	0.504466	0.004596	0.9964
	34158	0.754483	11.24499	0.0000
	56190	6.629492	4.474234	0.0015

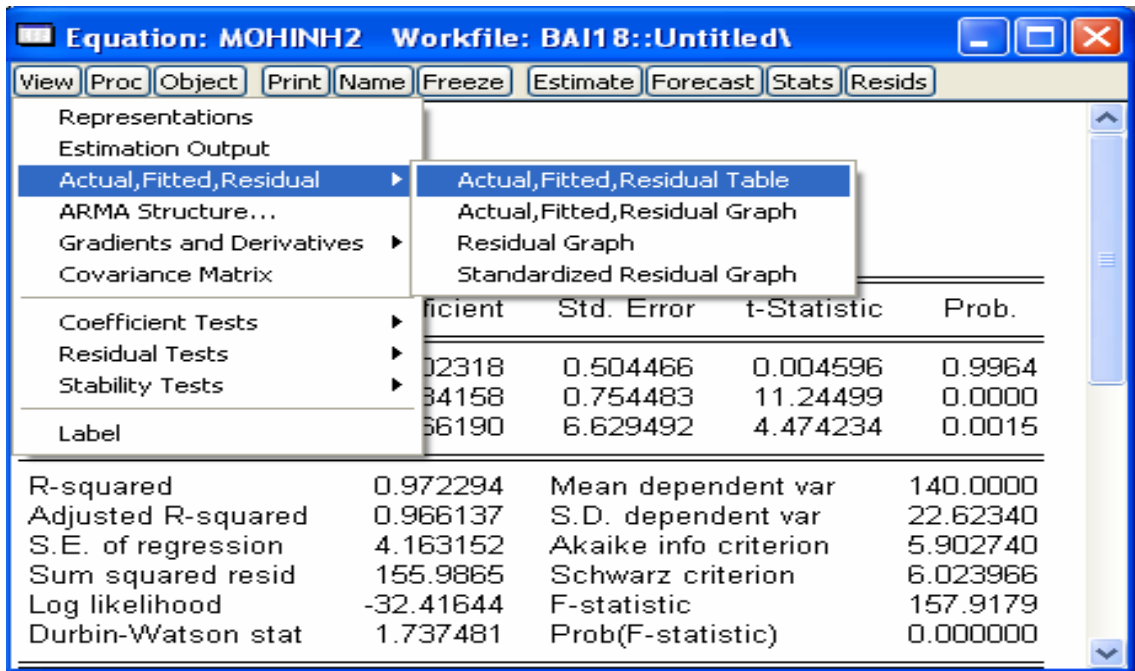
R-squared	0.972294	Mean dependent var	140.0000
Adjusted R-squared	0.966137	S.D. dependent var	22.62340
S.E. of regression	4.163152	Akaike info criterion	5.902740
Sum squared resid	155.9865	Schwarz criterion	6.023966
Log likelihood	-32.41644	F-statistic	157.9179
Durbin-Watson stat	1.737481	Prob(F-statistic)	0.000000

Nhấp chuột, kết quả tìm được là ma trận hiệp phương sai của các hệ số hồi quy như sau:

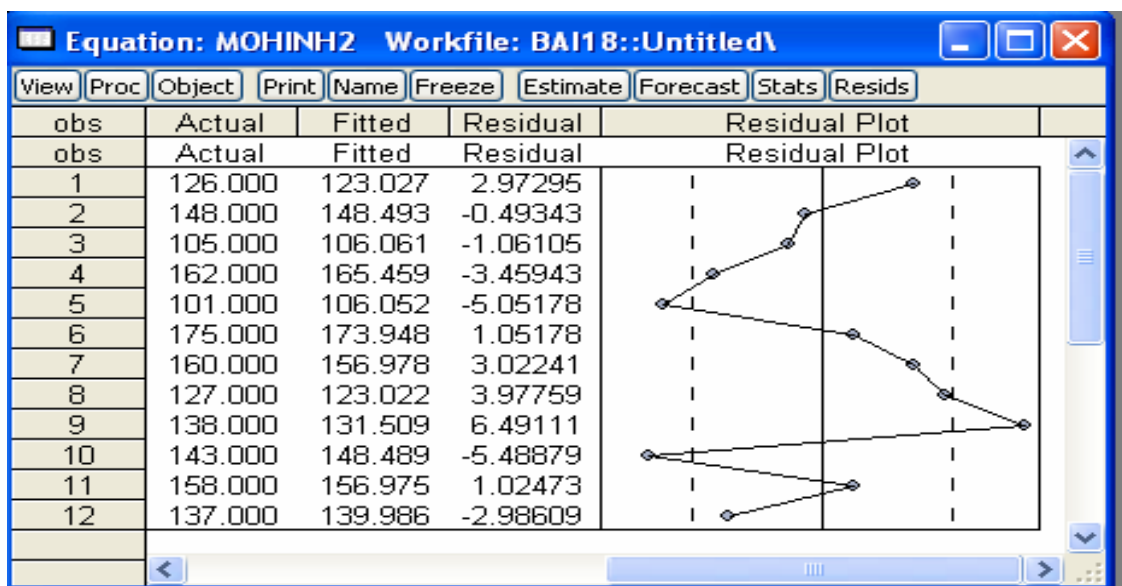


Coefficient Covariance Matrix				
	X2	X3	C	
X2	0.254486	-0.294668	-1.004550	
X3	-0.294668	0.569245	-1.801492	
C	-1.004550	-1.801492	43.95017	

Nếu bạn chọn **View/Actual, Fitted, Residual**, khi đó màn hình sẽ như sau:



Nhấp chuột từng lựa chọn một, các bạn sẽ thấy kết quả và từ đó rút ra bài học cho mình. Một kết quả tìm được như sau:



Bảng này cho ta biết giá trị quan sát thật ($Actual - Y_i$), giá trị ước lượng ($Fitted - \hat{Y}_i$), phần dư ($Residual - e_i$), và biểu đồ phần dư ($Residual Plot$).

Trên cửa sổ “**Equation**” bạn bấm vào thực đơn **View**, sau đó lựa chọn từng “**món**” có trong thực đơn để xem kết quả là gì nhé. Đừng sợ! Đây là cách học rất hiệu quả khi không có sự hướng dẫn của giáo viên hay các chuyên gia phần mềm Eviews.

Chương 3: KIỂM ĐỊNH CÁC GIẢ THUYẾT CỦA MÔ HÌNH HỒI QUY TUYẾN TÍNH CỔ ĐIỂN BẰNG EViews

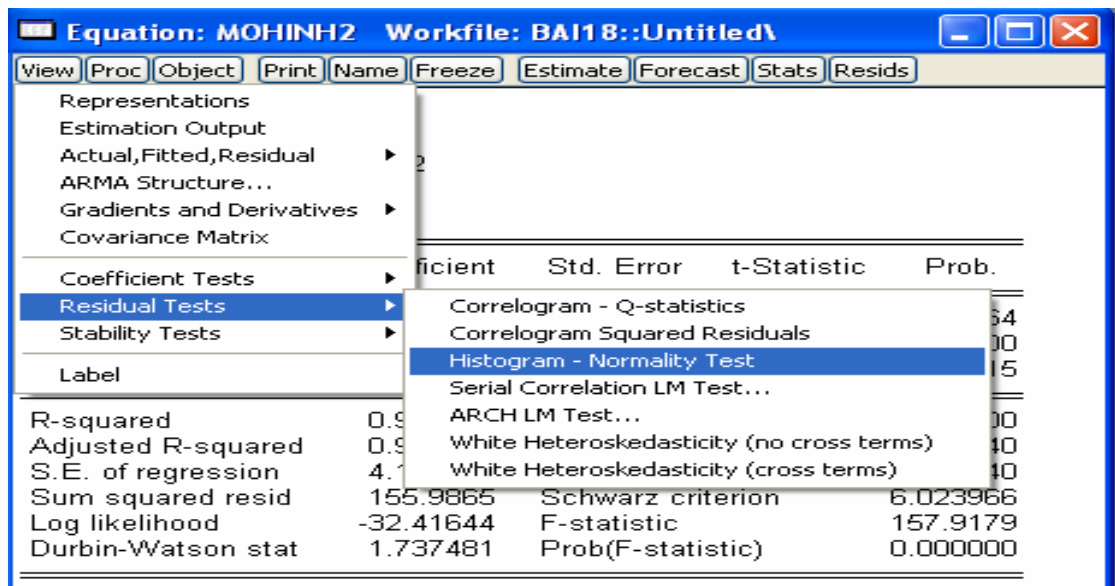
3.1. Kiểm định giả thuyết về phân phối chuẩn của sai số ngẫu nhiên U_i .

Giả thuyết: H_0 : U_i tuân theo phân phối chuẩn

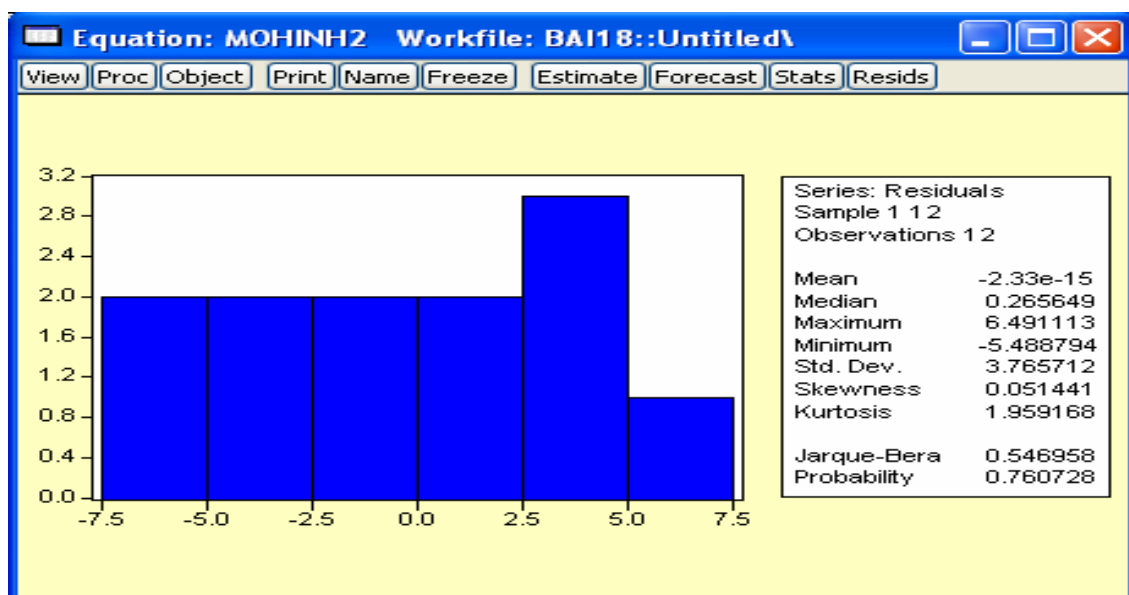
H_1 : U_i không tuân theo phân phối chuẩn.

Giả thuyết $U_i \sim N(0, \sigma^2)$ của mô hình hồi quy tuyến tính cổ điển có thể được kiểm định bằng thống kê **Jarque – Bera (JB)** như sau:

Từ cửa sổ Equation: MOHINH2 bạn chọn **View/Residual Tests/Histogram – Normality Test**, khi đó màn hình như sau:



Nhấp chuột, kết quả tìm được như sau:



Kết quả trên cho ta thấy $\mathbf{JB} = 0.5469$ và $\mathbf{P-value} = 0.7607$. Giá trị P-value này quá lớn nên ta không bác bỏ giả thuyết H_0 . Kết luận U_i có phân phối chuẩn.

3.2. Kiểm định giả thuyết về tính đa cộng tuyến.

Để biết có hiện tượng đa cộng tuyến hay không, bạn có thể sử dụng ma trận tương quan giữa các biến độc lập trong mô hình và xem xét độ mạnh giữa các biến. Cách thức tiến hành trên Eviews tôi đã giới thiệu ở phần trên.

Ngoài ra, bạn cũng có thể sử dụng một số phương pháp khác như: Mô hình hồi quy phụ, thừa số tăng phương sai (VIF),...

3.3. Kiểm định giả thuyết về phương sai thay đổi của sai số ngẫu nhiên U_i .

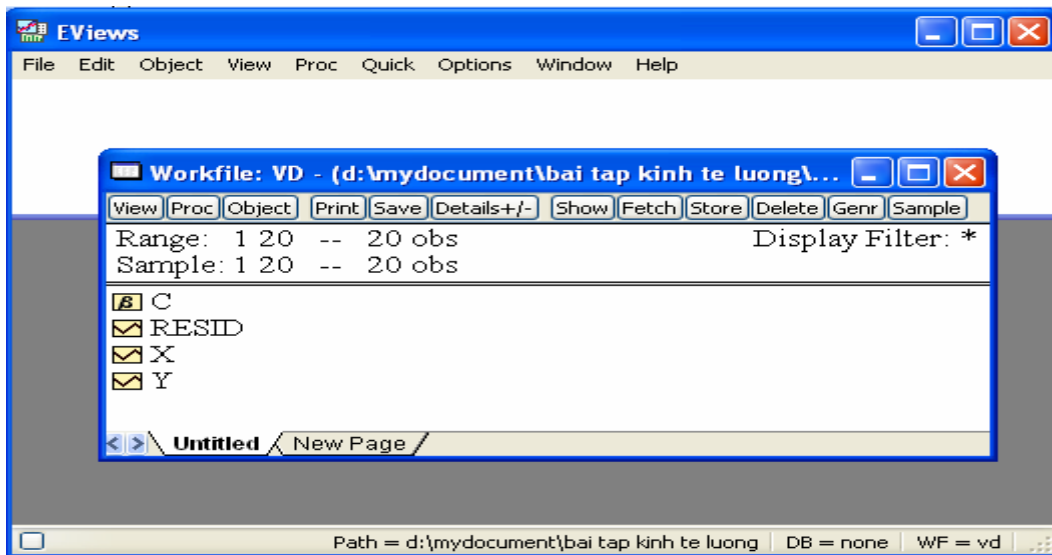
Để phát hiện phương sai của sai số có thay đổi không ta có nhiều cách thức khác nhau.

Ví dụ: Cho các số liệu quan sát về chi tiêu cho tiêu dùng (Y) và thu nhập (X) hàng tháng của 20 hộ gia đình ở một vùng nông thôn (đơn vị: 10.000đ).

Y	X	Y	X	Y	X
19,9	22,3	25,5	26,1	14,8	16,4
31,2	32,3	10,3	10,3	21,6	24,1
31,8	33,6	38,8	40,2	29,3	30,1
12,1	12,1	8,0	8,1	25,0	28,3
40,7	42,3	33,1	34,5	17,9	18,2
6,1	6,2	33,5	38,0	19,8	20,1
38,6	44,7	13,1	14,1		

- Vẽ đồ thị phần dư. Cho nhận xét?
- Kiểm định Park? Cho nhận xét?
- Kiểm định Glejser? Cho nhận xét?
- Kiểm định White (dạng có tích chéo và không có tích chéo).

Đầu tiên các bạn mở workfile này ra (nếu đã lưu từ trước) hoặc các bạn phải tạo ra file làm việc cho ví dụ này.



Cách 1: Phương pháp đồ thị.

Để vẽ đồ thị phân tán của phần dư theo biến độc lập X , trước hết bạn phải ước lượng hàm hồi quy Y theo X . Đặt tên cho hàm này là **EQ01**. Kết quả ước lượng như sau:

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.707476	0.659498	1.072750	0.2976
X	0.910260	0.023894	38.09502	0.0000

R-squared	0.987749	Mean dependent var	23.55500
Adjusted R-squared	0.987068	S.D. dependent var	10.78691
S.E. of regression	1.226673	Akaike info criterion	3.341128
Sum squared resid	27.08507	Schwarz criterion	3.440701
Log likelihood	-31.41128	F-statistic	1451.230
Durbin-Watson stat	2.320725	Prob(F-statistic)	0.000000

Sau khi ước lượng xong, bạn có thể tạo ra chuỗi giá trị cho Y ước lượng (\hat{Y}_i) bằng cách chọn **Forecast** trên cửa sổ **EQ01**. Một cửa sổ mới được tạo ra như sau:

Trong khung **Forecast name** bạn thấy máy tính đã tự động tạo ra một chuỗi dự báo có tên là **yf** (dĩ nhiên bạn có thể thay đổi tên này và đặt tên mới tùy thích). Đây chính là chuỗi giá trị Y ước lượng (\hat{Y}_i). Chọn **OK** là hoàn tất.

Hãy đóng cửa sổ **EQ01** lại. Giờ bạn hãy mở phần dư (resid) ra. Từ cửa sổ **Resid**, chọn **Object/Copy Object...** Lựa chọn này cho phép bạn copy ra một chuỗi phần dư mới. Từ chuỗi mới này bạn vào **Name** và đặt tên cho nó (ở đây tôi đặt tên cho nó là **RESIDEQ01**), và ghi nhớ đây là phần dư được tạo ra từ phương trình hồi quy gốc (**EQ01**). Sau này nếu chúng ta cần đến phần dư gốc để thực hiện một số kiểm định thì bạn chọn vào **RESIDEQ01** nhé, vì vector phần dư (**Resid**) do máy tạo ra nhằm để chứa phần dư của một mô hình vừa mới ước lượng sau cùng. Do đó, vector phần dư “**resid**” sẽ thay đổi khi ta ước lượng nhiều phương trình khác nhau. Giờ thì bạn hiểu tại sao tôi khuyên bạn nên copy ra một chuỗi phần dư mới và đặt tên cho nó rồi chứ!

Series: RESID Workfile: VD::Untitled\

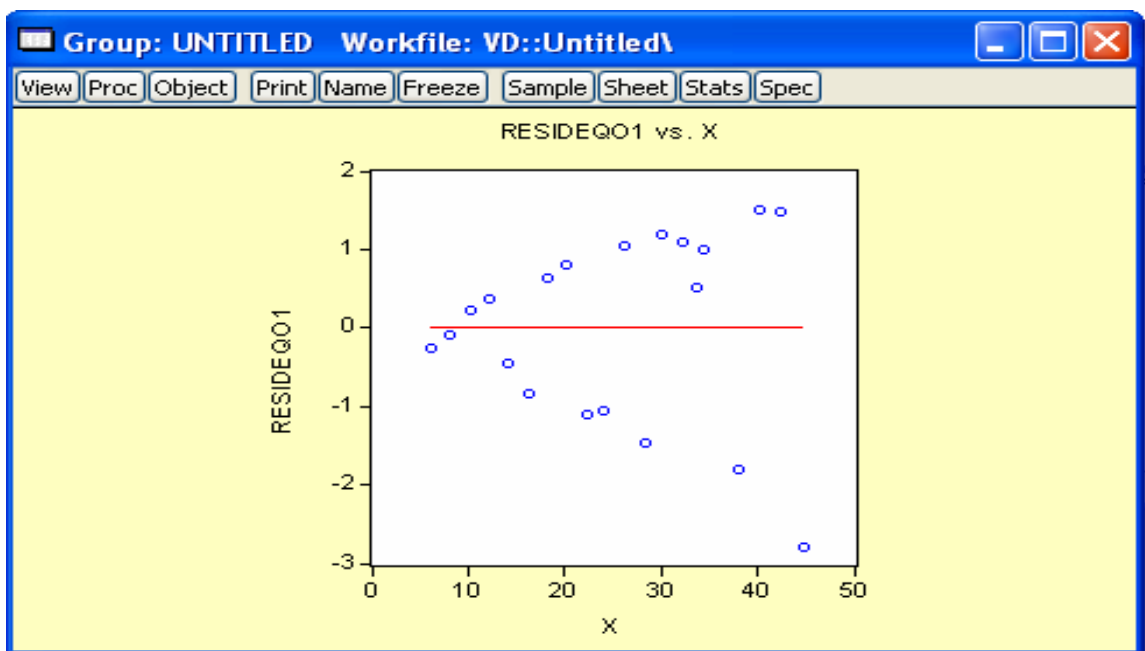
View Proc Object Properties Print Name Freeze Default Sort Edit+/- Smpl+/- Lat

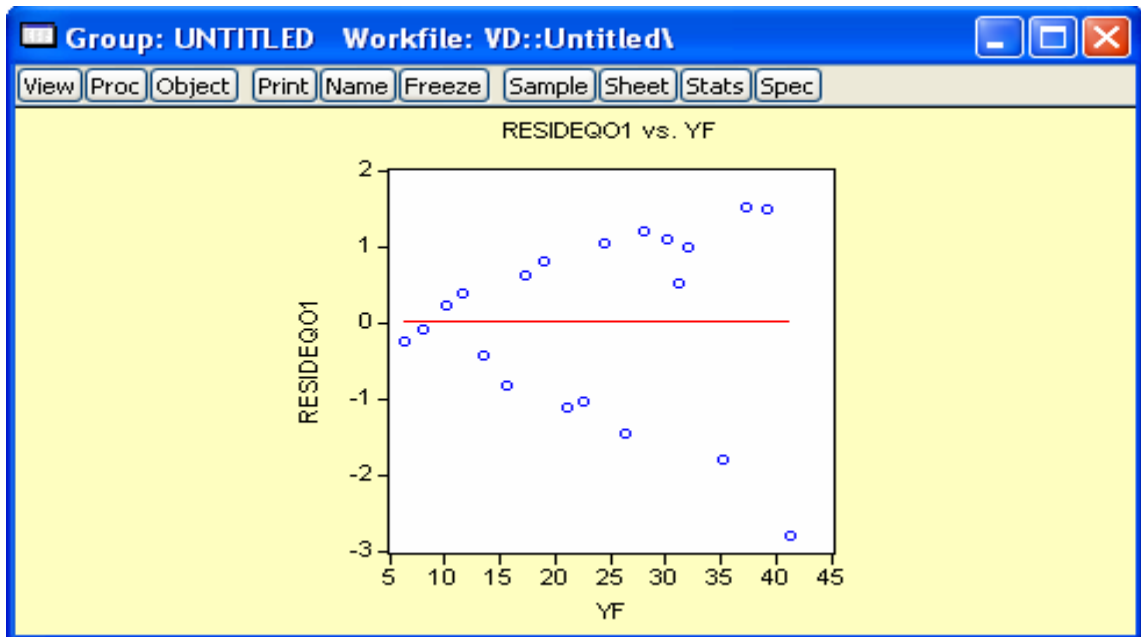
RESID

dated: 05/16/09 - 10:07

Observation	RESID
1	
2	
3	
4	
5	
6	-0.251088
7	-2.796094
8	1.034740
9	0.216847
10	1.500075
11	-0.080581
12	0.099557

Từ cửa sổ Workfile, bạn có thể vẽ đồ thị phân tán của phần dư theo biến độc lập X (cho hồi quy đơn), nếu là hồi quy bội bạn vẽ theo từng biến độc lập hoặc theo giá trị Y ước lượng (\hat{Y}_i). Đồ thị phần dư theo biến độc lập và theo giá trị \hat{Y}_i có dạng như sau:





Lưu ý: Trong hồi quy đơn nếu bạn vẽ đồ thị phần dư theo biến độc lập X hoặc theo \hat{Y}_i thì hình dạng của nó giống nhau.

Nhìn vào đồ thị trên ta thấy có hiện tượng phương sai của sai số thay đổi. Tuy nhiên, phương pháp đồ thị chỉ cho ta cái nhìn trực quan từ đó phán đoán xem có hiện tượng phương sai thay đổi hay không thôi, nó không thay thế được các kiểm định.

Cách 2: Phát hiện phương sai thay đổi bằng các kiểm định.

- Kiểm định Park:

Kiểm định Park đòi hỏi bạn phải ước lượng mô hình sau:

$$\ln(e_i^2) = \beta_1 + \beta_2 \ln(X_i) + v_i$$

Sau khi ước lượng mô hình trên, bạn phải đi kiểm định giả thuyết:

$$H_0: \beta_2 = 0 \rightarrow \text{Phương sai của sai số không đổi.}$$

$$H_1: \beta_2 \neq 0 \rightarrow \text{Phương sai của sai số thay đổi.}$$

Kết quả ước lượng mô hình trên như sau:

Equation: UNTITLED Workfile: VD::Untitled\				
View Proc Object Print Name Freeze Estimate Forecast Stats Resids				
Dependent Variable: LOG(RESIDEQ01^2)				
Method: Least Squares				
Date: 05/16/09 Time: 11:17				
Sample: 1 20				
Included observations: 20				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-8.529469	1.038973	-8.209520	0.0000
LOG(X)	2.581552	0.330972	7.799916	0.0000
R-squared	0.771686	Mean dependent var	-0.553733	
Adjusted R-squared	0.759002	S.D. dependent var	1.676648	
S.E. of regression	0.823093	Akaike info criterion	2.543145	
Sum squared resid	12.19468	Schwarz criterion	2.642718	
Log likelihood	-23.43145	F-statistic	60.83870	
Durbin-Watson stat	1.516201	Prob(F-statistic)	0.000000	

Kết quả kiểm định Park cho ta kết luận có hiện tượng phương sai của sai số thay đổi. Vì $P\text{-value}(\beta_2) = 0.0000$ quá nhỏ. Ta kết luận bác bỏ H_0 .

- Kiểm định Glejser:

Kiểm định Glejser đòi hỏi bạn phải ước lượng các mô hình sau:

$$|e_i| = \beta_1 + \beta_2 X_i + v_i \quad (1)$$

$$|e_i| = \beta_1 + \beta_2 \sqrt{X} + v_i \quad (2)$$

$$|e_i| = \beta_1 + \beta_2 1/X + v_i \quad (3)$$

$$|e_i| = \beta_1 + \beta_2 1/\sqrt{X} + v_i \quad (4)$$

Sau khi ước lượng các mô hình trên, bạn phải đi kiểm định giả thuyết:

$$H_0: \beta_2 = 0 \rightarrow \text{Phương sai của sai số không đổi.}$$

$$H_1: \beta_2 \neq 0 \rightarrow \text{Phương sai của sai số thay đổi.}$$

Kết quả ước lượng các mô hình trên như sau:

Mô hình 1:

Equation: UNTITLED Workfile: VD::Untitled\

View Proc Object Print Name Freeze Estimate Forecast Stats Resids

Dependent Variable: ABS(RESIDEQ01)
 Method: Least Squares
 Date: 05/16/09 Time: 14:12
 Sample: 1 20
 Included observations: 20

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-0.166276	0.191084	-0.870170	0.3957
X	0.045755	0.006923	6.608966	0.0000

R-squared 0.708164 Mean dependent var 0.982184
 Adjusted R-squared 0.691951 S.D. dependent var 0.640368
 S.E. of regression 0.355419 Akaike info criterion 0.863598
 Sum squared resid 2.273802 Schwarz criterion 0.963171
 Log likelihood -6.635981 F-statistic 43.67843
 Durbin-Watson stat 1.781207 Prob(F-statistic) 0.000003

Mô hình 2:

Equation: UNTITLED Workfile: VD::Untitled\

View Proc Object Print Name Freeze Estimate Forecast Stats Resids

Dependent Variable: ABS(RESIDEQ01)
 Method: Least Squares
 Date: 05/16/09 Time: 14:14
 Sample: 1 20
 Included observations: 20

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-1.070759	0.338661	-3.161740	0.0054
SQR(X)	0.422536	0.067597	6.250797	0.0000

R-squared 0.684611 Mean dependent var 0.982184
 Adjusted R-squared 0.667090 S.D. dependent var 0.640368
 S.E. of regression 0.369482 Akaike info criterion 0.941211
 Sum squared resid 2.457307 Schwarz criterion 1.040784
 Log likelihood -7.412107 F-statistic 39.07246
 Durbin-Watson stat 1.809645 Prob(F-statistic) 0.000007

Mô hình 3:

Equation: UNTITLED Workfile: VD::Untitled\

View Proc Object Print Name Freeze Estimate Forecast Stats Resids

Dependent Variable: ABS(RESIDEQ01)
 Method: Least Squares
 Date: 05/16/09 Time: 14:16
 Sample: 1 20
 Included observations: 20

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	1.647269	0.187054	8.806400	0.0000
1/X	-12.30281	2.880094	-4.271668	0.0005

R-squared	0.503409	Mean dependent var	0.982184
Adjusted R-squared	0.475821	S.D. dependent var	0.640368
S.E. of regression	0.463628	Akaike info criterion	1.395172
Sum squared resid	3.869120	Schwarz criterion	1.494745
Log likelihood	-11.95172	F-statistic	18.24715
Durbin-Watson stat	1.610693	Prob(F-statistic)	0.000459

Mô hình 4:

Equation: UNTITLED Workfile: VD::Untitled\

View Proc Object Print Name Freeze Estimate Forecast Stats Resids

Dependent Variable: ABS(RESIDEQ01)
 Method: Least Squares
 Date: 05/16/09 Time: 14:17
 Sample: 1 20
 Included observations: 20

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	2.532172	0.325954	7.768500	0.0000
1/SQR(X)	-6.972226	1.401908	-4.973384	0.0001

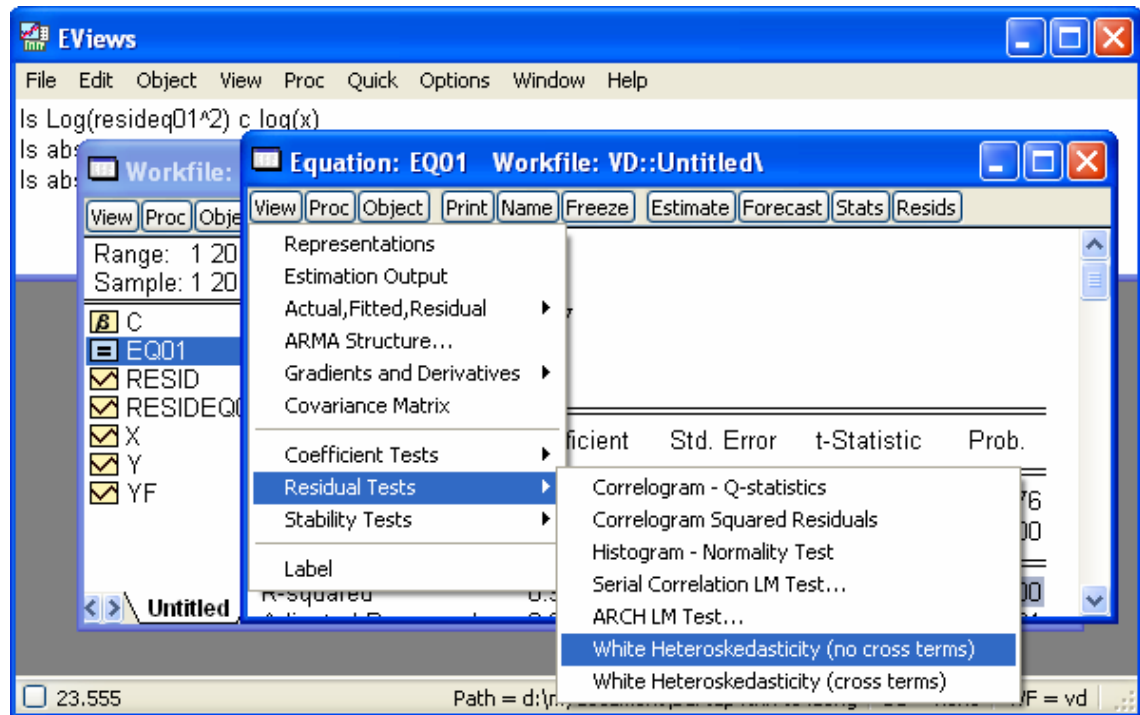
R-squared	0.578795	Mean dependent var	0.982184
Adjusted R-squared	0.555395	S.D. dependent var	0.640368
S.E. of regression	0.426989	Akaike info criterion	1.230525
Sum squared resid	3.281761	Schwarz criterion	1.330098
Log likelihood	-10.30525	F-statistic	24.73455
Durbin-Watson stat	1.659035	Prob(F-statistic)	0.000098

Kết quả ước lượng 4 dạng mô hình mà Glejser đề xuất đều cho kết quả bác bỏ giả thuyết H_0 . Có nghĩa là có hiện tượng phương sai của sai số thay đổi.

- Kiểm định White:

Kiểm định White đã được lập trình sẵn trên Eviews, do đó bạn không phải mất thời gian để ước lượng như kiểm định Park và Glejser. Để thực hiện kiểm định này bạn làm như sau:

Từ cửa sổ Workfile bạn mở phương trình hồi quy gốc (EQ01) ra. Trên cửa sổ EQ01 bạn chọn **View/Residual Tests/White Heteroskedasticity (no cross terms)** dạng không có tích chéo và **View/Residual Tests/White Heteroskedasticity (cross terms)** dạng có tích chéo.



Kết quả kiểm định White (dạng không có tích chéo) như sau:

The screenshot shows the results of the White Heteroskedasticity Test in the 'Equation: EQ01' window. The test statistics and test equation are displayed.

White Heteroskedasticity Test:			
F-statistic	14.45799	Prob. F(2,17)	0.000215
Obs*R-squared	12.59517	Prob. Chi-Square(2)	0.001841

Test Equation:
 Dependent Variable: RESID^2
 Method: Least Squares
 Date: 05/16/09 Time: 14:32
 Sample: 1 20
 Included observations: 20

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	1.224820	1.215234	1.007888	0.3276
X	-0.147350	0.107503	-1.370656	0.1883
X^2	0.005025	0.002086	2.408643	0.0276

Nhìn vào kết quả trên cho ta thấy: $\chi^2 = nR^2 = 12.59517$ (Obs*R-squared = 12.59517) và P-value = 0.00841 (Prob. Chi-Square(2) = 0.001841). Kết luận bác bỏ H_0 , có nghĩa là có phương sai thay đổi.

Ngoài các kiểm định trên, còn có một số kiểm định khác nữa như kiểm định Nhân Tử Larrange (LM). Đặt trung tổng quát mô hình hồi quy bội là:

$$Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + \dots + \beta_k X_{ki} + U_i$$

Các phương trình hồi quy phụ sau đây đã bao quát hầu hết các trường hợp giả thuyết về nguyên nhân gây ra phương sai của sai số thay đổi.

- **Breusch-Pagan:** $\sigma_i^2 = \alpha_1 + \alpha_2 Z_{2i} + \alpha_3 Z_{3i} + \dots + \alpha_p Z_{pi} + v_i$

- **Glejser:** $|\sigma_i| = \alpha_1 + \alpha_2 Z_{2i} + \alpha_3 Z_{3i} + \dots + \alpha_p Z_{pi} + v_i$

- **Harvey-Godfrey:** $\ln(\sigma_i^2) = \alpha_1 + \alpha_2 Z_{2i} + \alpha_3 Z_{3i} + \dots + \alpha_p Z_{pi} + v_i$

Tuy nhiên, trong phạm vi tài liệu này các kiểm định trên không được giới thiệu ở đây. Bạn đọc có thể tự thực hiện các kiểm định trên theo trình tự các bước mà chúng ta đã học trên lớp.

3.4. Kiểm định giả thuyết về tự tương quan của sai số ngẫu nhiên U_i .

☛ Để phát hiện có tương quan chuỗi giữa các nhiễu hay không, ta có thể tiến hành theo một số cách sau:

Cách 1: Phương pháp đồ thị.

- **Cách 2:** Kiểm định Durbin Watson:

- **Cách 3:** Kiểm định nhân tử Larrange (LM) – Kiểm định BG.

- **Cách 4:** Sử dụng giản đồ tương quan (correlogram).

Ví dụ: File DATA-02.wf1 sau đây chứa dữ liệu về Chi Phí Bảo Trì Xe (COST) và Tuổi Thọ Xe (AGE). Sử dụng Eviews để:

a. Thực hiện một mô hình hồi qui đơn giản xác định rằng COST là một hàm tuyến tính theo AGE. Viết phương trình hồi qui mẫu của mô hình này. Sử dụng kiểm định dựa trên giá trị P-value, hãy xác định xem có mối quan hệ tuyến tính có ý nghĩa về mặt thống kê với mức ý nghĩa 5% giữa hai biến COST và AGE hay không.

b. Vẽ đồ thị phần dư của mô hình hồi qui ở câu a theo AGE. Dựa vào đồ thị nói trên. Anh/Chị có ý kiến gì về vấn đề tương quan chuỗi.

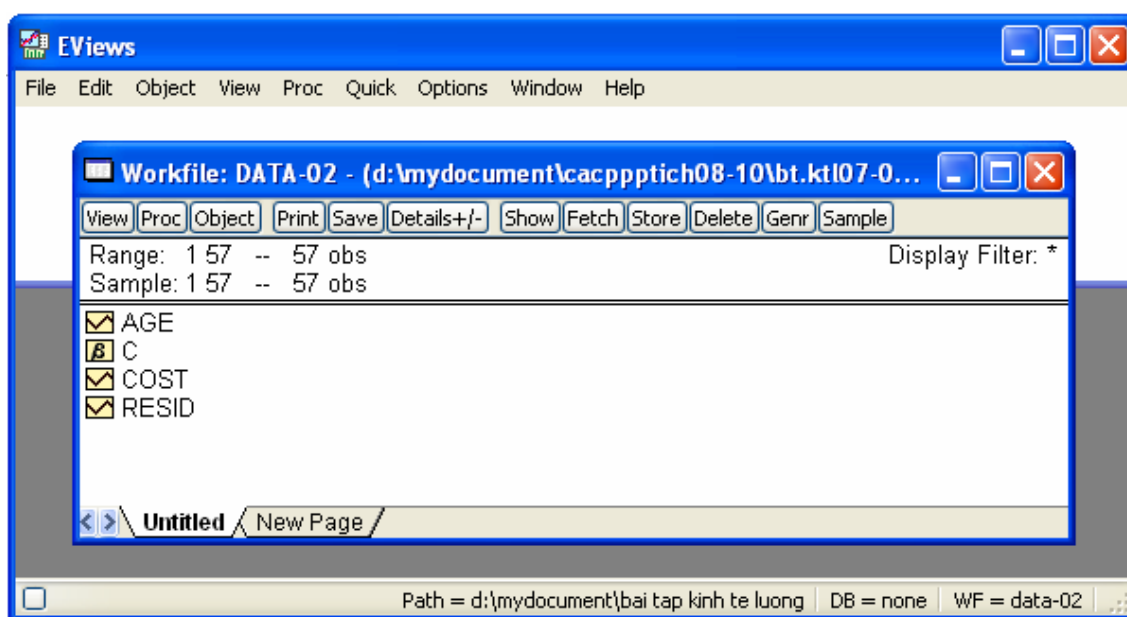
c. Tất nhiên, với dữ liệu chuỗi thời gian chúng ta nghi ngờ rằng thành phần sai số có thể có tương quan chuỗi bậc 1. Hãy trình bày kiểm định Durbin-Watson xem phần dư của phương trình hồi qui nói trên có tương quan chuỗi bậc 1 hay không?

d. Một sinh viên khóa 49 đề nghị nên dựa vào giản đồ tương quan (correlogram) để kiểm định phần dư có tương quan chuỗi bậc 1 hay không với mức ý nghĩa 5%. Anh/Chị hãy dựa vào giản đồ tương quan của phần dư của phép hồi qui nói trên để xác nhận sự tồn tại của tương quan chuỗi bậc 1.

e. Một sinh viên khóa 50 đề nghị nên dùng kiểm định nhân tử Lagrange để kiểm định phần dư có tương quan chuỗi bậc 1 hay không với mức ý nghĩa 5%. Anh/Chị hãy tiến hành kiểm định nhân tử Lagrange để xác nhận sự tồn tại của tương quan chuỗi bậc 1.

f. Hãy giải thích các hậu quả của việc tồn tại tương quan chuỗi đối với hàm ước lượng OLS. Hãy đề nghị giải pháp khắc phục tương quan chuỗi (nếu có) trong phép hồi qui nói trên.

File dữ liệu trên sau khi mở ra có dạng như sau:



a) Trước tiên bạn phải ước lượng mô hình COST theo AGE. Kết quả ước lượng cho ở bảng sau:

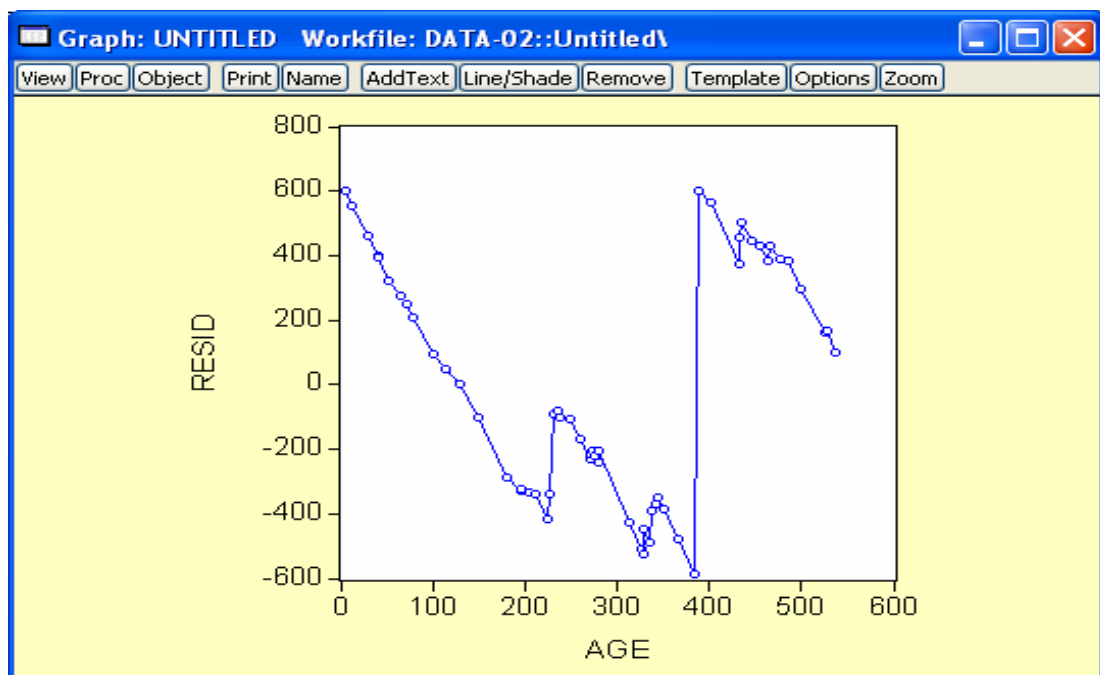
Equation: EQ01 Workfile: DATA-02::Untitled\				
View Proc Object Print Name Freeze Estimate Forecast Stats Resids				
Dependent Variable: COST				
Method: Least Squares				
Date: 05/19/09 Time: 21:55				
Sample: 1 57				
Included observations: 57				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
AGE	7.343478	0.329580	22.28136	0.0000
C	-625.9350	104.1496	-6.009962	0.0000
R-squared	0.900265	Mean dependent var	1426.632	
Adjusted R-squared	0.898451	S.D. dependent var	1151.186	
S.E. of regression	366.8453	Akaike info criterion	14.68222	
Sum squared resid	7401653.	Schwarz criterion	14.75390	
Log likelihood	-416.4431	F-statistic	496.4588	
Durbin-Watson stat	0.237982	Prob(F-statistic)	0.000000	

Bạn hãy lưu phương trình này với tên EQ01. Kết quả trên cho thấy AGE có ảnh hưởng đến COST vì P-value = 0.0000 < 5%.

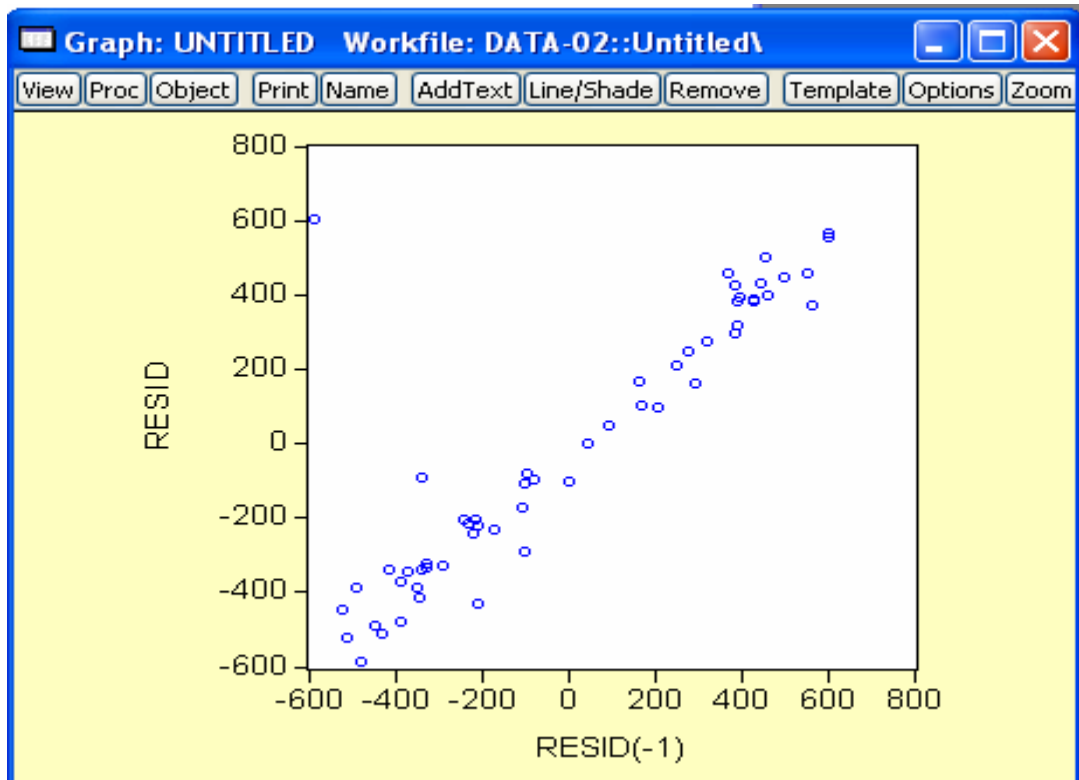
b) Phương pháp đồ thị.

Sau khi ước lượng mô hình, bạn có thể vẽ đồ thị phần dư để nhận định có tự tương quan hay không. Ở trên tôi đã hướng dẫn các bạn cách vẽ đồ thị, do vậy việc vẽ các đồ thị sau đây không phải là quá khó đối với bạn!

- Vẽ đồ thị phần dư e_t (hoặc e_t^2) theo AGE.

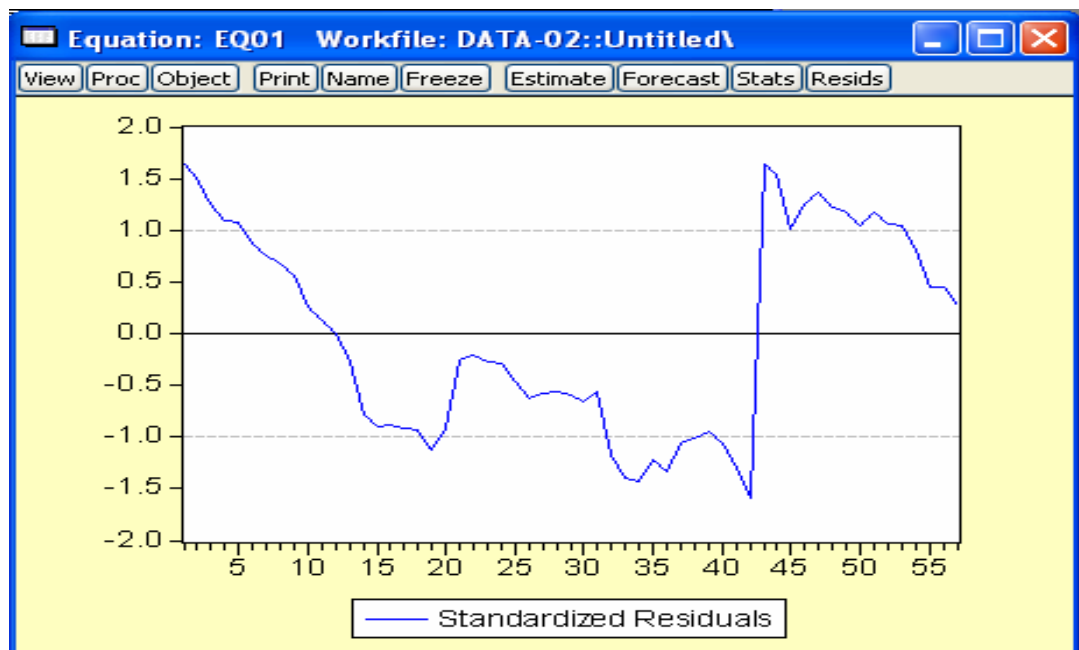


- Vẽ đồ thị phần dư e_t theo e_{t-1} (còn gọi là lược đồ tự tương quan bậc nhất - AR(1))



- Vẽ đồ thị phần dư chuẩn hóa ($\frac{e_t}{\sigma}$) theo thời gian.

Từ cửa sổ EQ01, bạn chọn **View/Actual/Fitted/Residual/Standardized Residual Graph**. Khi đó, đồ thị phần dư chuẩn hóa như sau:



Các đồ thị trên đều cho kết quả có tự tương quan giữa các nhiễu U_i .

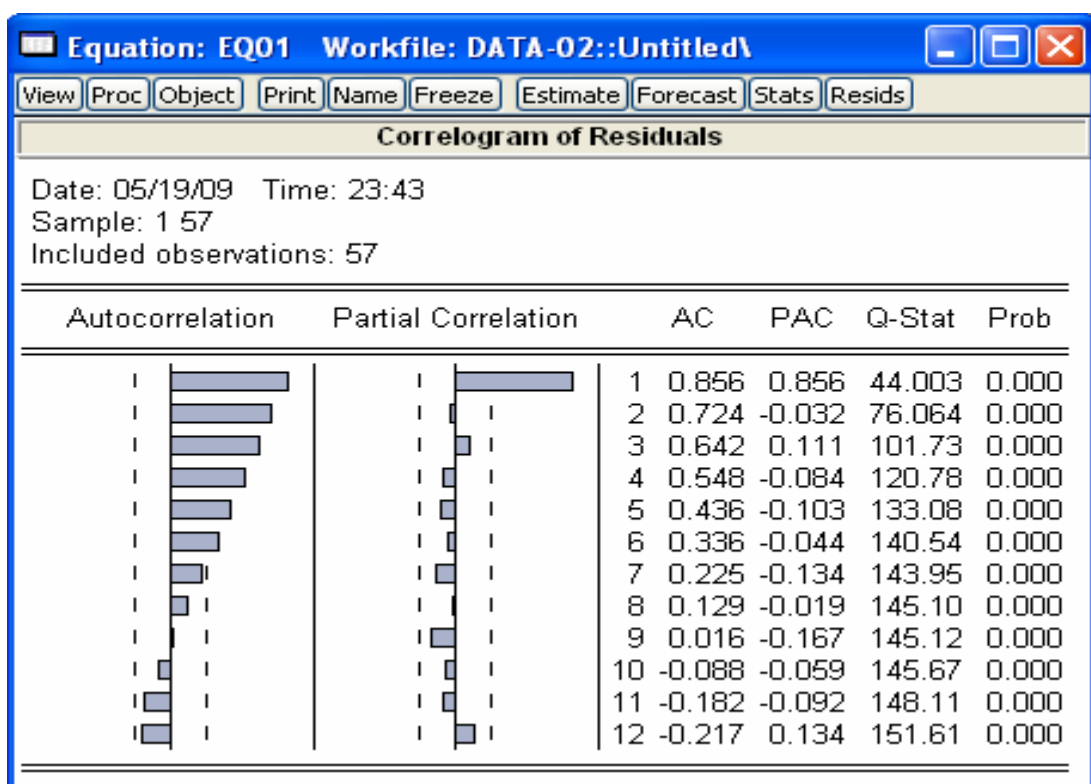
c) Kiểm định Durbin – Watson cho tương quan chuỗi bậc 1:

Kết quả ước lượng cho ta thấy thống kê Durbin – Watson là:

Durbin-Watson stat	0.237982
--------------------	----------

Ta thấy, $d = 0.237982 < 1$. Kết luận có tự tương quan dương. (Áp dụng theo quy tắt kinh nghiệm).

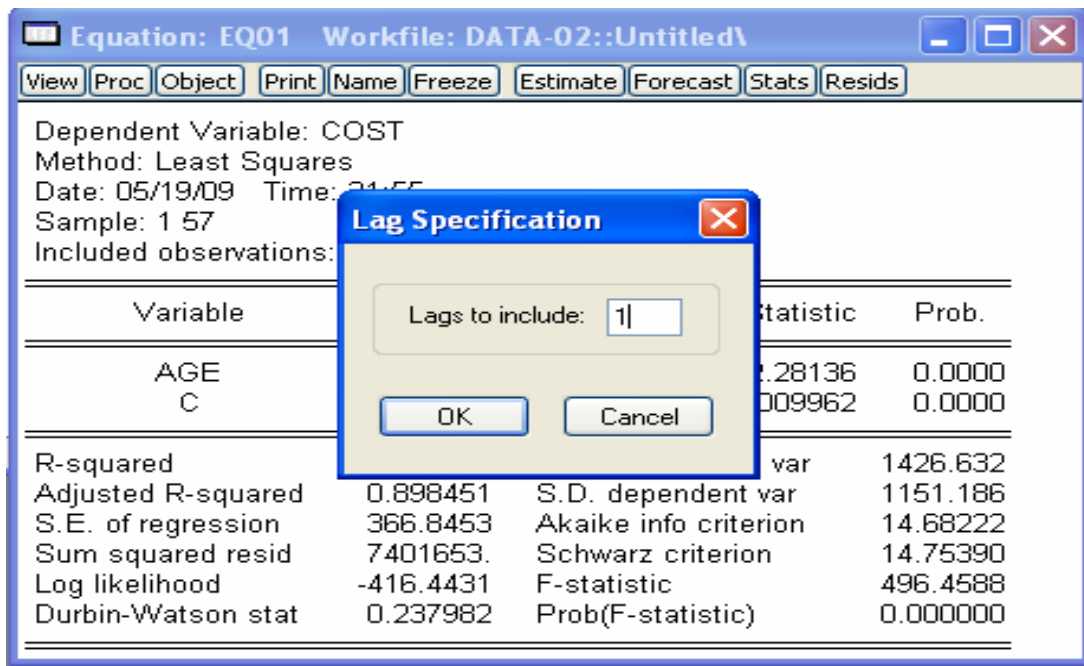
d) Sử dụng giản đồ tương quan (correlogram). Từ cửa sổ EQ01, bạn chọn **View/Residual Tests/Correlogram – Q - Statistics**, kết quả cho ở bảng sau:



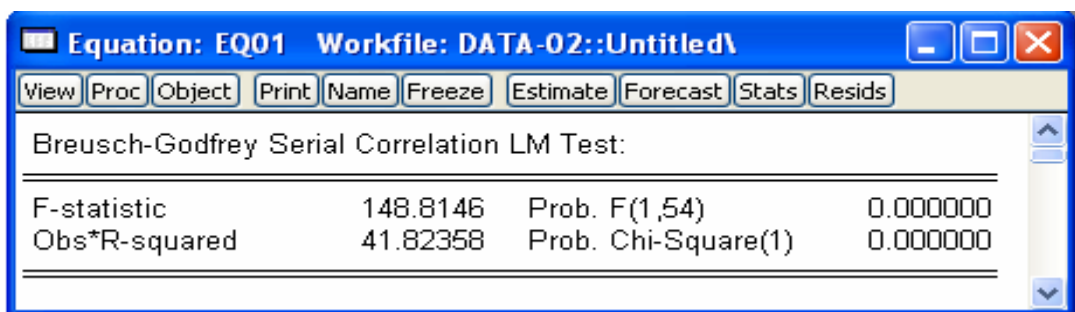
Kết luận có tương quan chuỗi vì $P\text{-value} = 0.000 < 5\%$ ứng với tự tương quan bậc 1. Nên bác bỏ giả thuyết cho rằng không có tương quan chuỗi. Các bậc tự tương quan từ bậc 2 đến 12 đều cho kết quả tương tự.

e) Kiểm định nhân tử Larrange (LM) – Kiểm định BG.

Từ cửa sổ EQ01, bạn chọn **View/Residual Tests/Serial Correlation LM Test...** Một màn hình hiện ra như sau.



Trong khung “**Lags to include**” nhập vào số 1 (Nếu muốn kiểm tra tự tương quan bậc 2 thì nhập số 2, bậc 3 nhập số 3,...). Bấm OK, kết quả như sau:



Kết luận có tương quan chuỗi vì thống kê χ^2 và thống kê F đều cho kết quả P-value = 0.00000 < 5%. Nên bác bỏ giả thuyết cho rằng không có tương quan chuỗi.

☛ Khắc phục tự tương quan:

Có nhiều cách để khắc phục hiện tượng tự tương quan. Ở đây, chỉ giới thiệu cách khắc phục bằng phương pháp lặp **Cochrane – Orcutt** trên Eviews (vì Eviews đã lập trình sẵn phương pháp này). Các phương pháp khác không được lập trình sẵn nên bạn đọc thực hiện một cách thủ công theo trình tự các bước chúng ta đã thảo luận trên lớp, và nên làm trên Eviews sẽ nhanh hơn. Các bạn có thể xem phần **help** trong Eviews về cách khắc phục theo phương pháp Cochrane – Orcutt.

Estimating AR Models

First-Order Serial Correlation

To estimate an AR(1) model in EViews, open an equation by selecting **Quick/Estimate Equation...** and enter your specification as usual, adding the special expression "AR(1)" to the end of your list. For example, to estimate a simple consumption function with AR(1) errors,

$$CS_t = c_1 + c_2GDP_t + u_t$$

$$(17.5) \quad u_t = \rho u_{t-1} + \epsilon_t$$

you should specify your equation as:

```
cs c gdp ar(1)
```

EViews automatically adjusts your sample to account for the lagged data used in estimation, estimates the model, and reports the adjusted sample along with the remainder of the estimation output.

Higher-Order Serial Correlation

Estimating higher order AR models is only slightly more complicated. To estimate an AR(*k*), you should enter your specification, followed by expressions for each AR term you wish to include. If you wish to estimate a model with autocorrelations from one to five:

$$CS_t = c_1 + c_2GDP_t + u_t$$

$$(17.6) \quad u_t = \rho_1 u_{t-1} + \rho_2 u_{t-2} + \dots + \rho_5 u_{t-5} + \epsilon_t$$

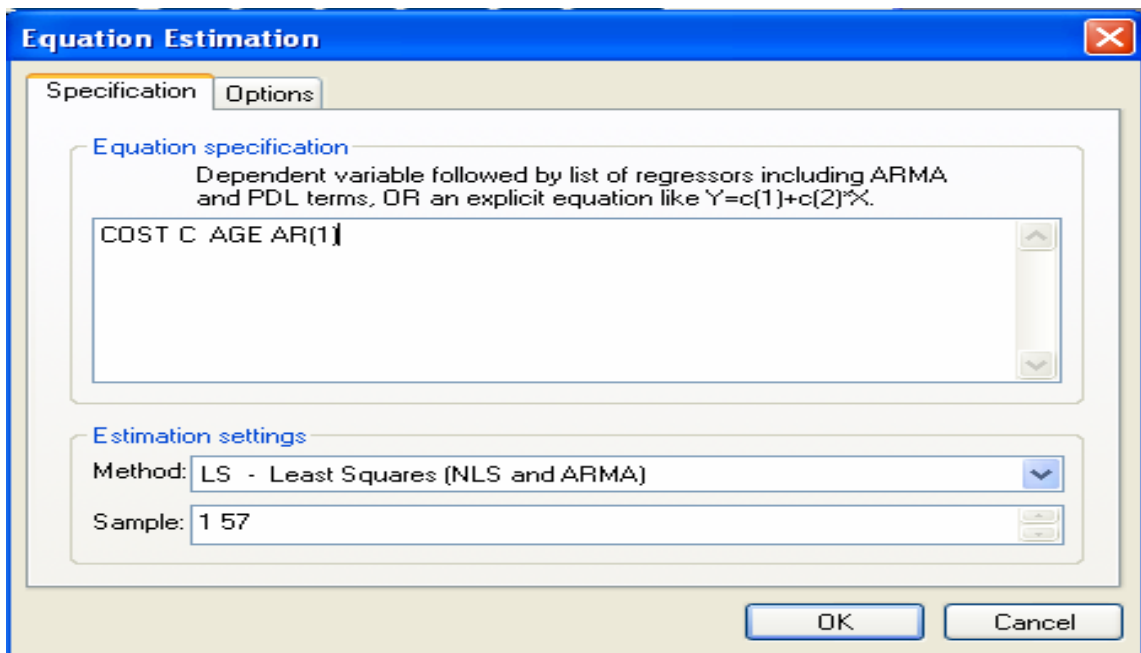
you should enter:

```
cs c gdp ar(1) ar(2) ar(3) ar(4) ar(5)
```

By requiring that you enter all of the autocorrelations you wish to include in your model, EViews allows you great flexibility in restricting lower order correlations to be zero. For example, if you have quarterly data and want to include a single term to account for seasonal autocorrelation, you could enter

```
cs c gdp ar(4)
```

Để sử dụng phương pháp lặp **Cochrane – Orcutt**, ta quay lại **thủ tục ước lượng mô hình gốc**. Ở vế phải của mô hình ta khai báo bậc của tự tương quan bằng cách đánh máy vào AR(*k*), với *k* là bậc của tự tương quan. Lưu ý là bạn cần phải khai báo một cách đầy đủ bậc của tự tương quan nhé. Chẳng hạn, nếu bậc của tự tương quan là 1, 2, và 4 thì các bạn phải đánh vào bên phải của mô hình là AR(1), AR(2), AR(4). Những số hạng này cách nhau ít nhất một khoảng trống. Ở ví dụ trên, bậc của tự tương quan là bậc 1 nên ta khai báo là AR(1). Các bạn có thể nhìn thấy cách làm ở hình sau:



Sau khi khai báo xong, bạn chọn OK, kết quả sẽ như sau:

Equation: UNTITLED Workfile: DATA-02::Untitled\

View Proc Object Print Name Freeze Estimate Forecast Stats Resids

Dependent Variable: COST
Method: Least Squares
Date: 05/23/09 Time: 16:46
Sample (adjusted): 2 57
Included observations: 56 after adjustments
Convergence achieved after 12 iterations

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-670.4550	411.1213	-1.630796	0.1089
AGE	7.283446	1.105175	6.590312	0.0000
AR(1)	0.858890	0.064338	13.34968	0.0000

R-squared	0.977750	Mean dependent var	1451.911
Adjusted R-squared	0.976910	S.D. dependent var	1145.529
S.E. of regression	174.0663	Akaike info criterion	13.20883
Sum squared resid	1605851.	Schwarz criterion	13.31733
Log likelihood	-366.8473	F-statistic	1164.511
Durbin-Watson stat	1.971430	Prob(F-statistic)	0.000000

Inverted AR Roots	.86
-------------------	-----

Kết quả ước lượng ở trên được hiểu là kết quả ước lượng ở bước cuối cùng của phương pháp lặp Cochrane – Orcutt. Ta được hàm hồi quy sau:

$$\text{COST} = -670.4549779 + 7.283446361 \cdot \text{AGE} + [\text{AR}(1)=0.8588899523]$$

$$\text{AR}(1)=0.8588899523 \text{ có nghĩa là } e_t = 0.8588899523e_{t-1} + \varepsilon_t$$

Các thông tin khác trong bảng kết quả trên được hiểu một cách bình thường.

TÀI LIỆU THAM KHẢO

1. Appendix A, EViews 5.1 Command and Programming Reference, Quantitative Micro Software.
2. Eviews 5.1 User's Guide, Quantitative Micro Software.
3. Nguyễn Quang Dong (2002) *Bài tập kinh tế lượng với sự trợ giúp của phần mềm Eviews*. Hà Nội: NXB Khoa học và Kỹ thuật.
4. Phạm Trí Cao - Vũ Minh Châu (2006) *Kinh Tế lượng ứng dụng*. TP. HCM: NXB Lao động Xã hội.
5. Ramu Ramanathan (2002) *Nhập Môn Kinh Tế Lượng*. NXB Harcourt, (Bản dịch tiếng Việt của Chương trình giảng dạy kinh tế Fulbright).

Chương 4. PHÂN TÍCH NHẬN DẠNG VÀ LỰA CHỌN MÔ HÌNH.

Chương 5: ỨNG DỤNG PHÂN TÍCH HỒI QUY – VẤN ĐỀ DỰ BÁO

GIỚI THIỆU SƠ LƯỢC PHÂN TÍCH DỮ LIỆU VỚI SPSS

TS. LÊ VĂN HUY
Email: levanhuy@vnn.vn

NỘI DUNG CHÍNH

- Giới thiệu về SPSS
- Mã hóa dữ liệu
- Giao diện và nhập liệu
- Một số điểm thường sử dụng

SƠ LƯỢC VỀ SPSS

- SPSS là gì?

SPSS FOR WINDOWS (Statistical Package for Social Sciences)

- Tại sao là SPSS?

- ⊕ STATA

- ⊕ EIEWS

- ⊕ EXCEL

- ⊕ SAS

MÃ HÓA DỮ LIỆU (CODING)

- Các loại mã hóa

 - ⊕ Mã hóa trước

 - ⊕ Mã hóa sau

- Ví dụ

 - ⊕ 1: Nam

 - 0: Nữ

 - ⊕ 1: Đồng ý

 - 0: Không đồng ý

 - ⊕ 1: Rất không đồng ý

 - 7: Rất đồng ý

- Chú ý khi Coding

MÃ HÓA DỮ LIỆU

- Các loại mã hóa

- ⊕ Mã hóa trước

- ⊕ Mã hóa sau

- Ví dụ

- ⊕ 1: Nam

- 0: Nữ

- ⊕ 1: Đồng ý

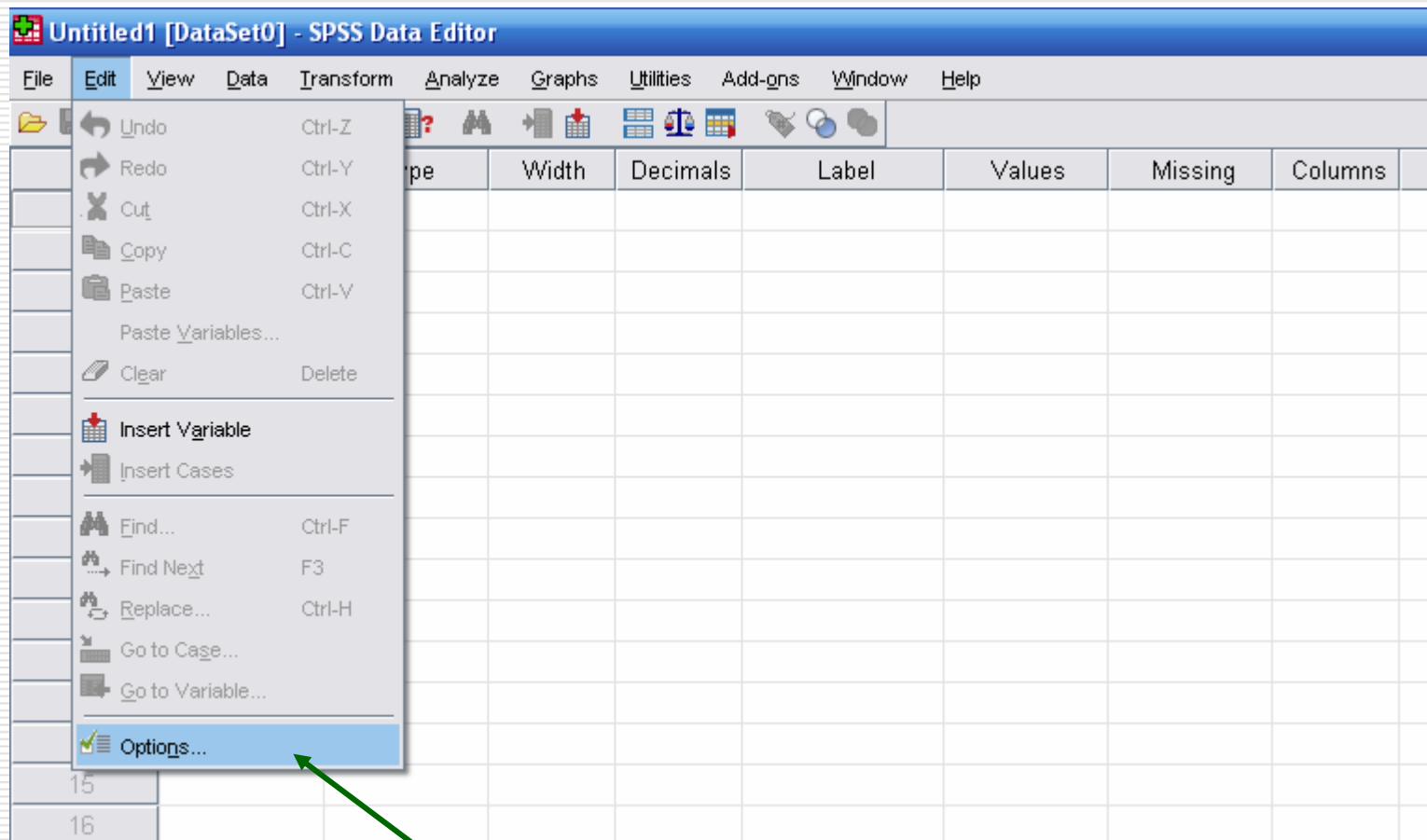
- 0: Không đồng ý

- ⊕ 1: Rất không đồng ý

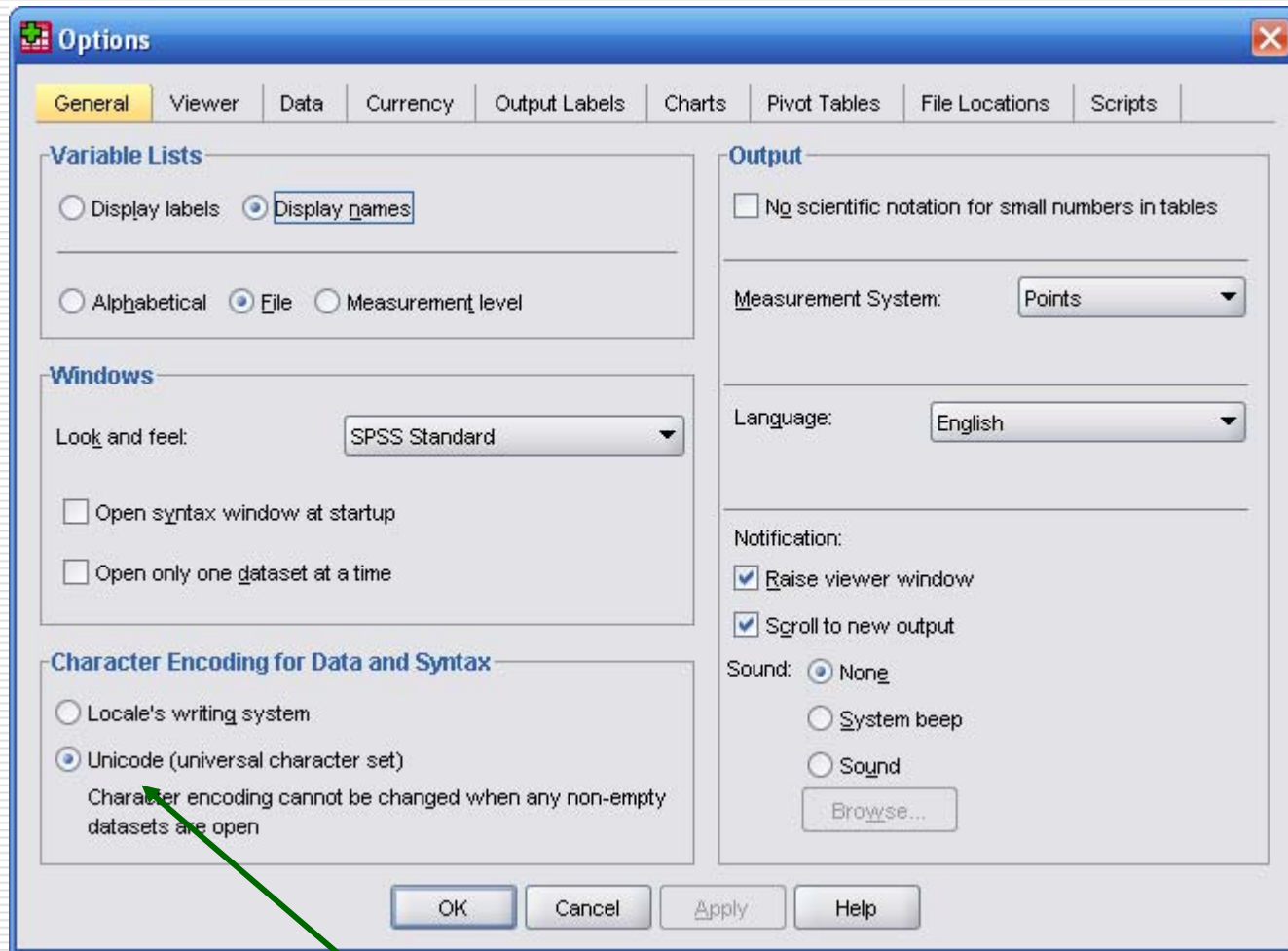
- 7: Rất đồng ý

- Chú ý khi Coding

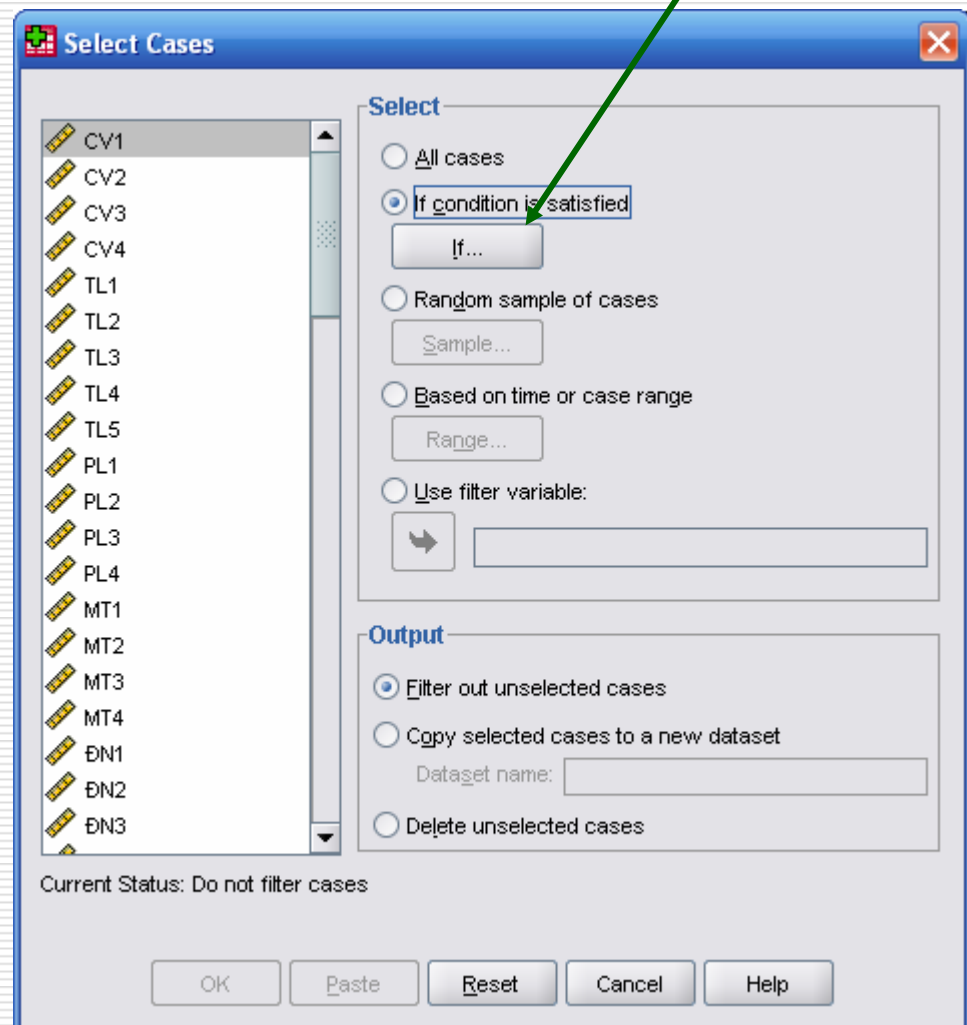
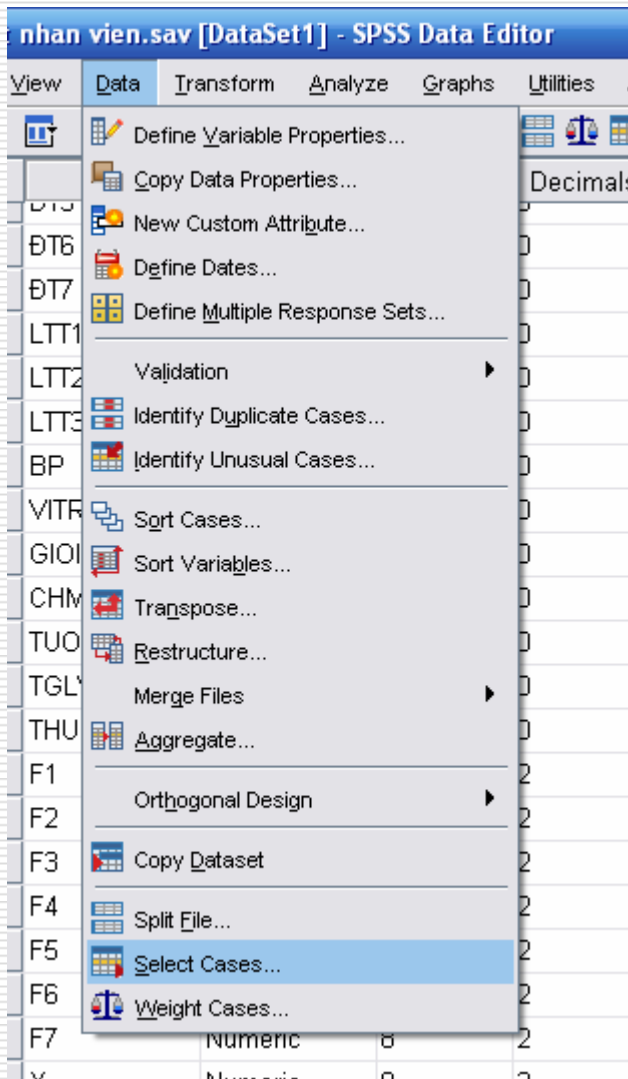
TIẾNG VIỆT (THEO UNICODE)



TIẾNG VIỆT (THEO UNICODE)

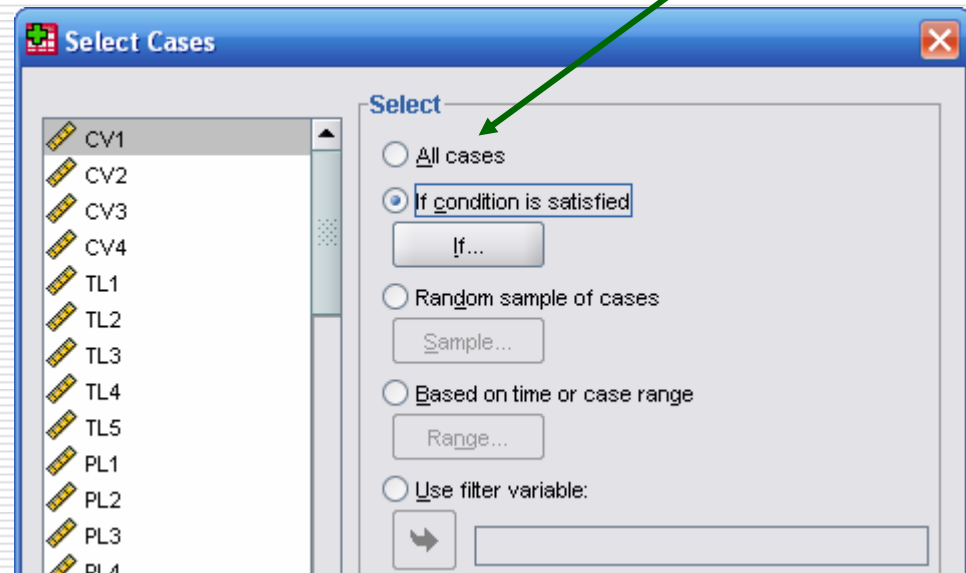


CHỌN DỮ LIỆU THEO ĐIỀU KIỆN

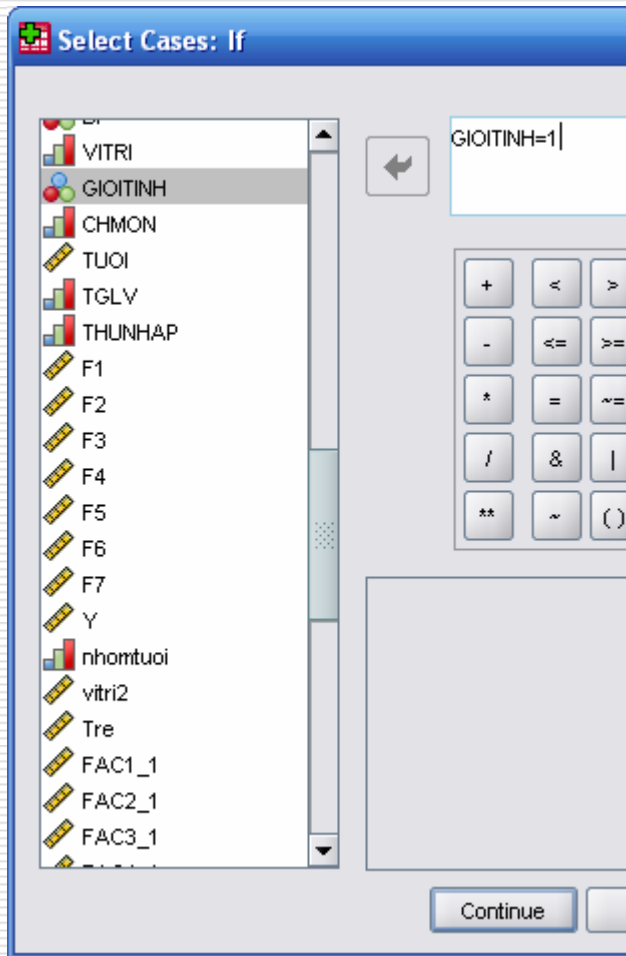


HÃY THỰC HÀNH

- Lựa chọn dữ liệu với điều kiện
 - ⊕ Giới tính = Nam
- Chú ý quan trọng: Sau khi chọn và phân tích sau, hãy trả về dữ liệu gốc khi chọn **All case**



CHỌN DỮ LIỆU THEO ĐIỀU KIỆN



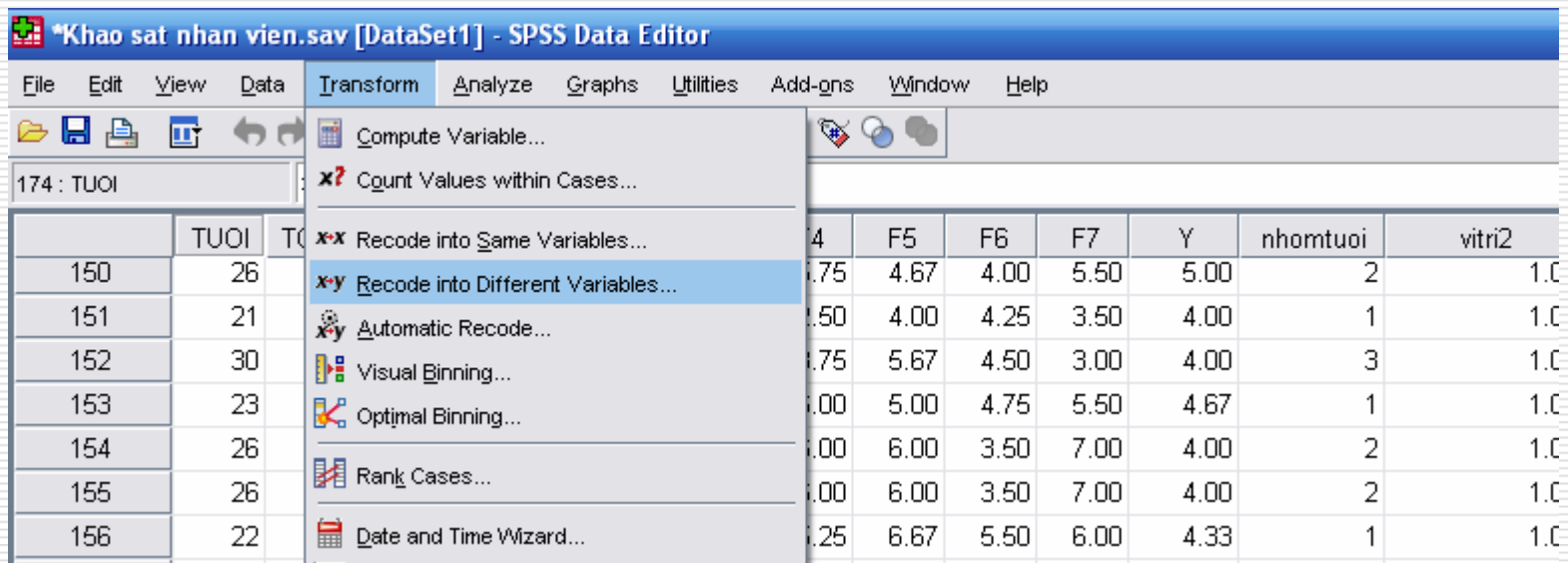
*Khao sat nhan vien.sav [DataSet1] - SPSS Data Editor

1 : CV1 4

	CV1	CV2	CV3	CV4	TL1	TL2
1	4	4	4	4	4	
2	4	4	4	3	3	
3	3	3	4	2	2	
4	5	6	5	5	4	
5	5	6	6	6	4	
6	7	7	4	7	4	
7	5	3	4	2	4	
8	5	4	5	6	4	
9	4	4	4	4	3	
10	6	6	7	4	3	
11	4	3	4	4	5	
12	3	3	5	3	3	
13	5	4	6	6	5	
14	6	5	4	7	5	
15	4	1	2	1	1	
16	3	3	2	5	5	
17	4	5	5	6	2	
18	5	5	6	3	4	

MÃ HÓA LẠI BIẾN

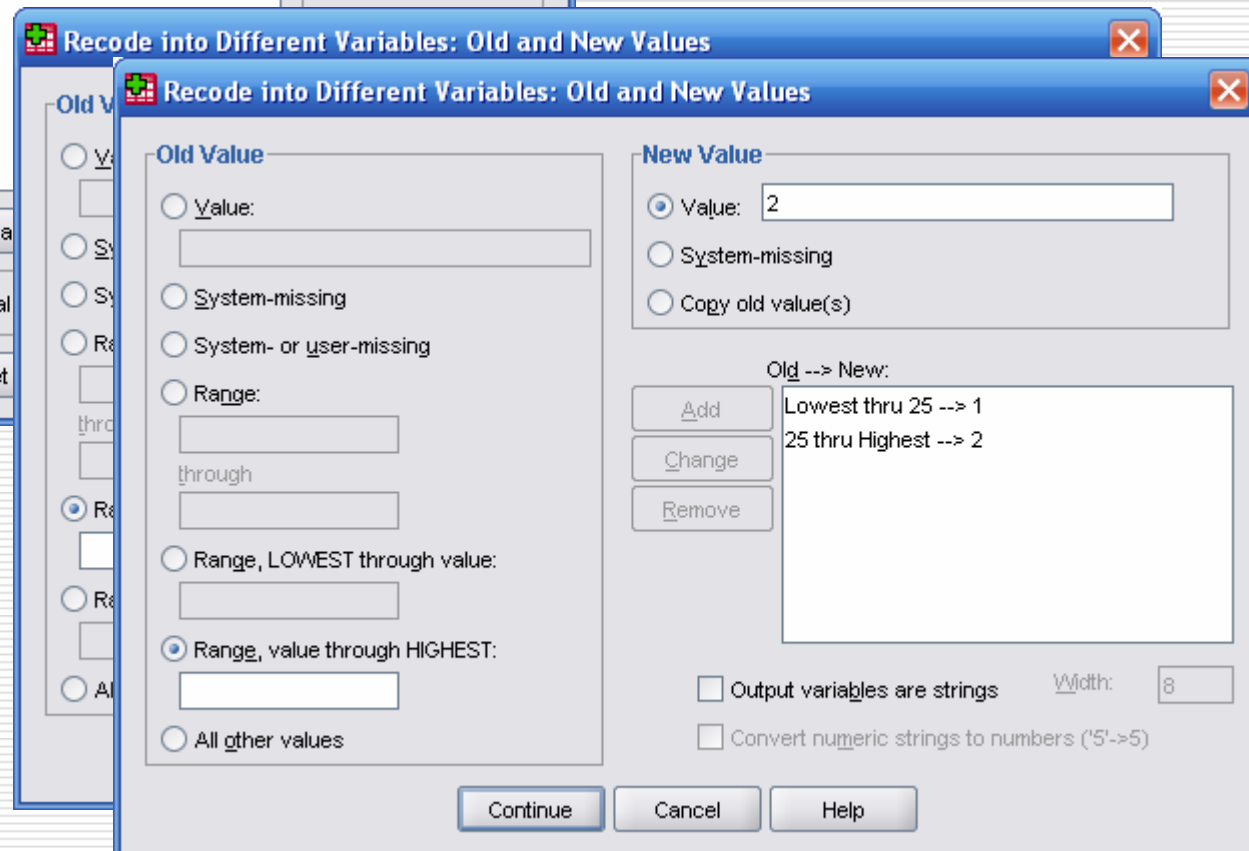
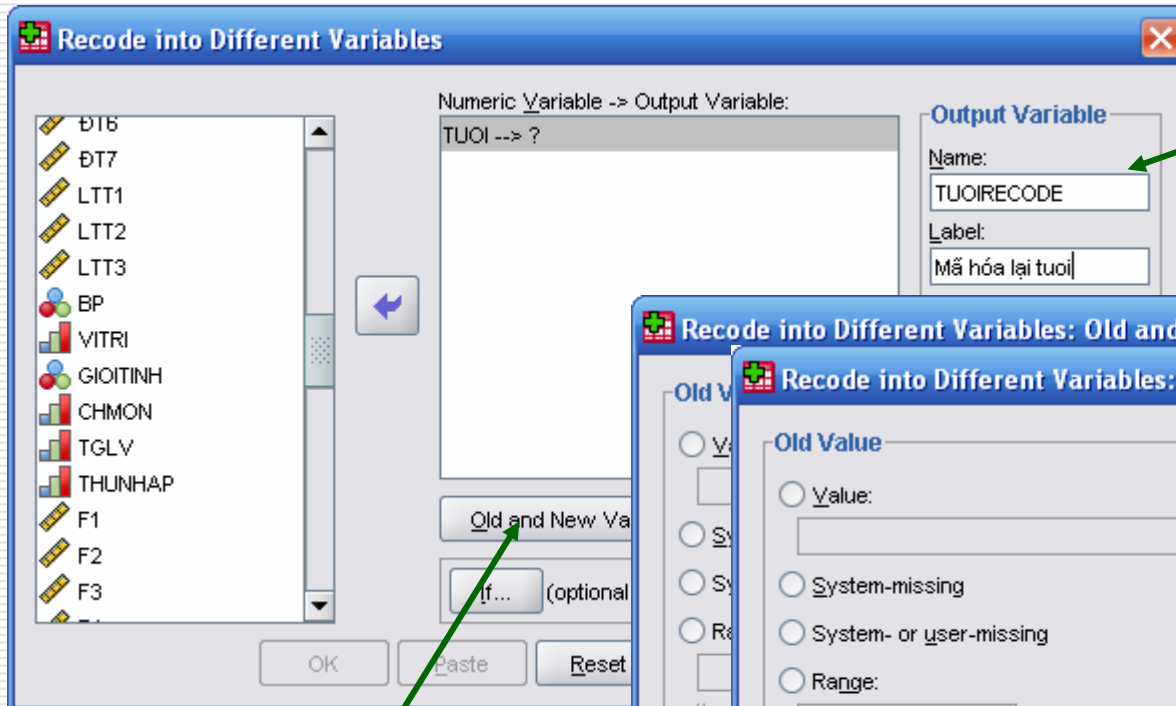
- Hãy mã hóa lại biến TUOI
 - ⊕ Tuổi < 25
 - ⊕ Tuổi > = 25
- Nên mã hóa và lưu lại trên **biến mới**



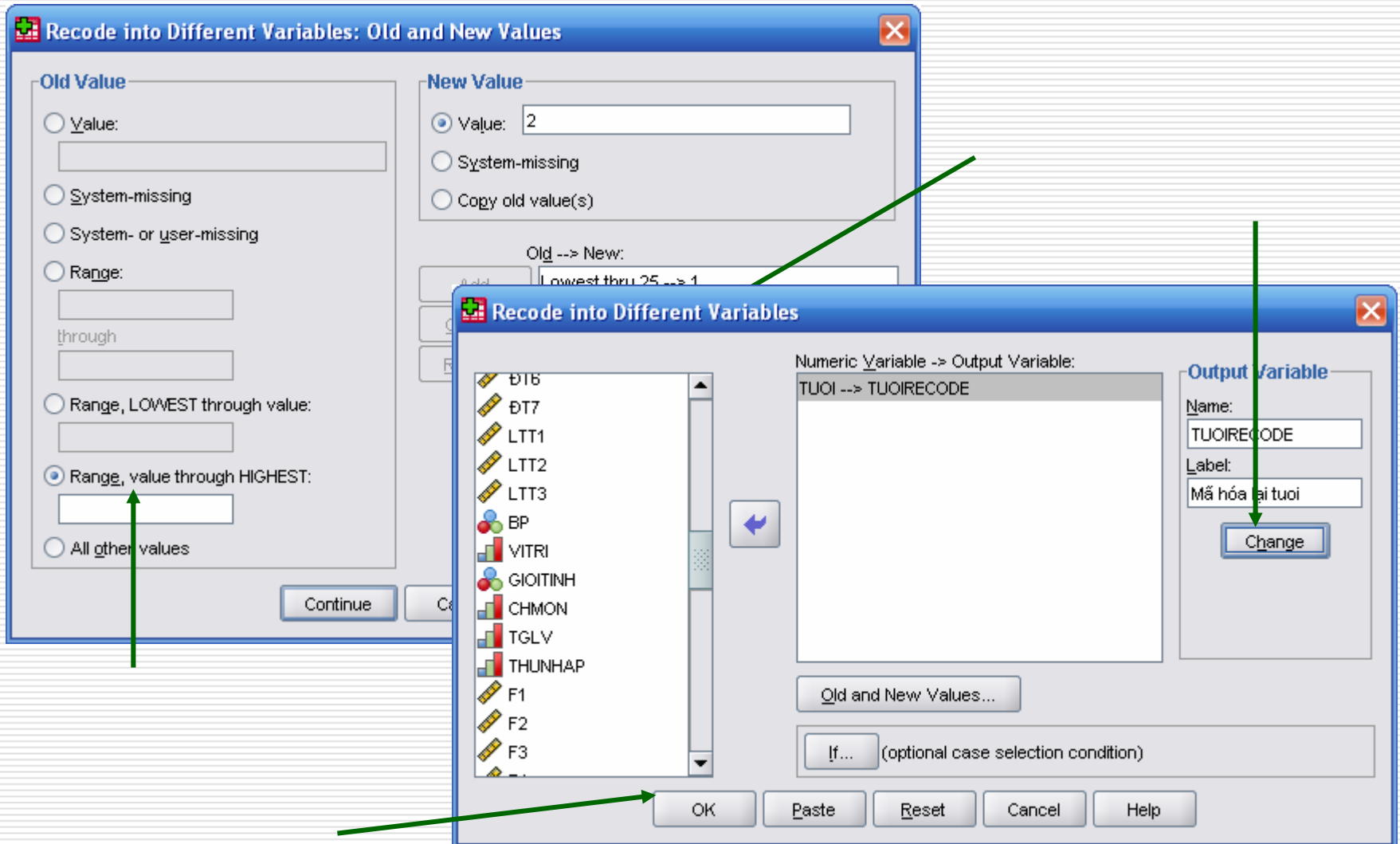
The screenshot shows the SPSS Data Editor interface with the 'Transform' menu open and the 'Recode into Different Variables...' option selected. The dialog box is open for the variable 'TUOI'. The main data grid shows the following values for 'TUOI': 26, 21, 30, 23, 26, 26, 22.

TUOI	TUOI	TUOI	F5	F6	F7	Y	nhomtuoi	vitri2	
150	26	26	4.75	4.67	4.00	5.50	5.00	2	1.0
151	21	21	4.50	4.00	4.25	3.50	4.00	1	1.0
152	30	30	4.75	5.67	4.50	3.00	4.00	3	1.0
153	23	23	5.00	5.00	4.75	5.50	4.67	1	1.0
154	26	26	5.00	6.00	3.50	7.00	4.00	2	1.0
155	26	26	5.00	6.00	3.50	7.00	4.00	2	1.0
156	22	22	5.25	6.67	5.50	6.00	4.33	1	1.0

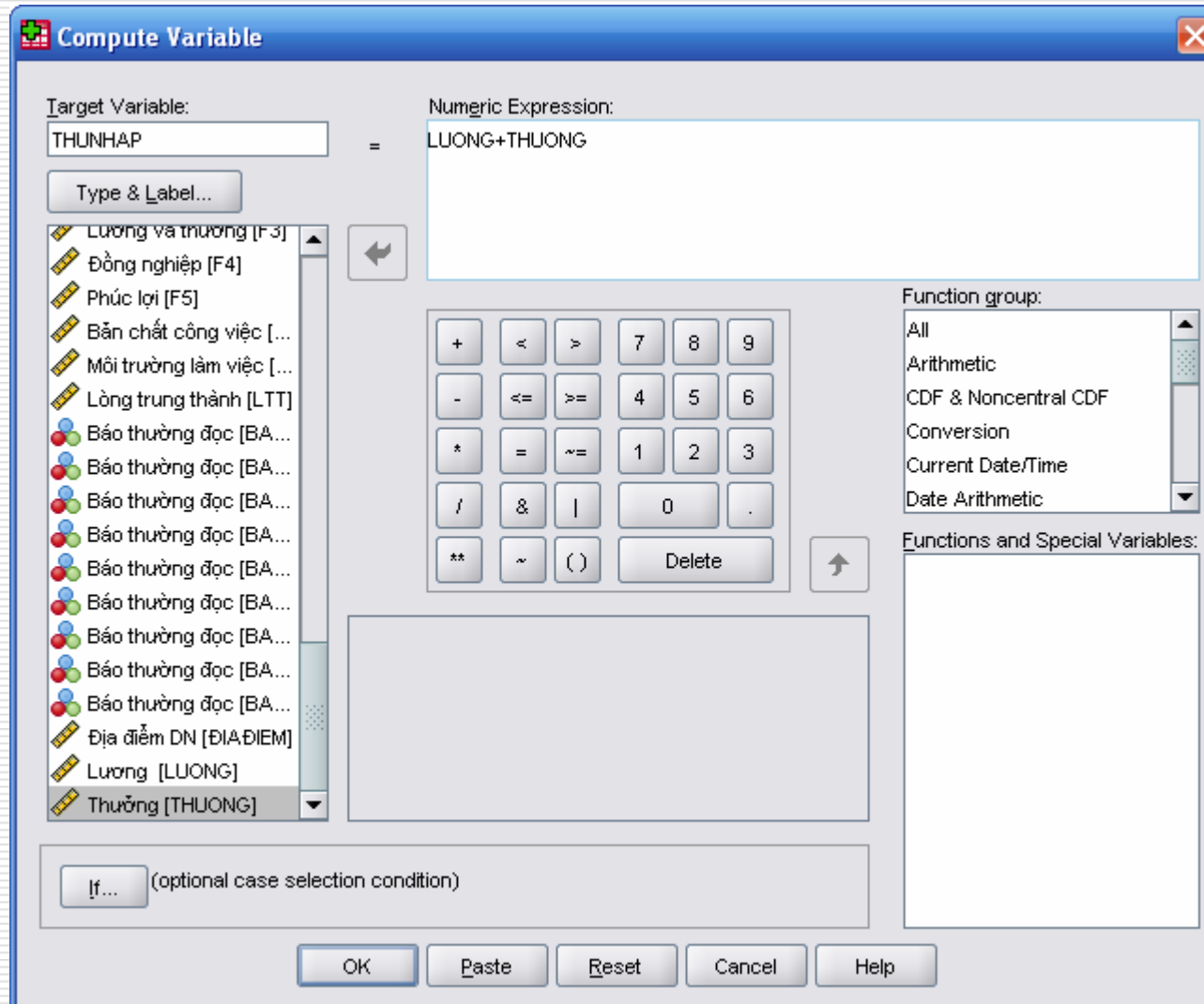
MÃ HÓA LẠI BIẾN



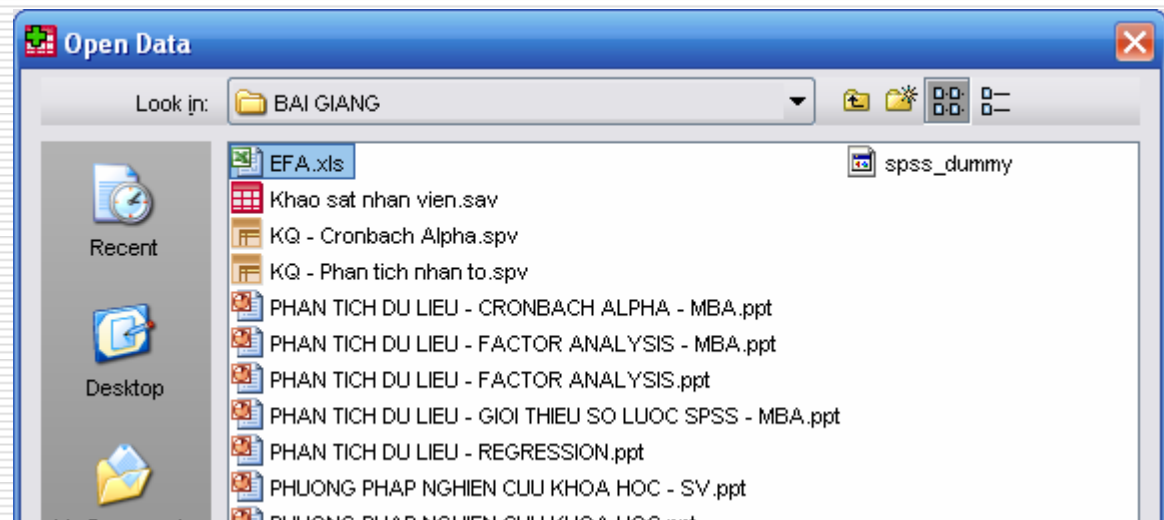
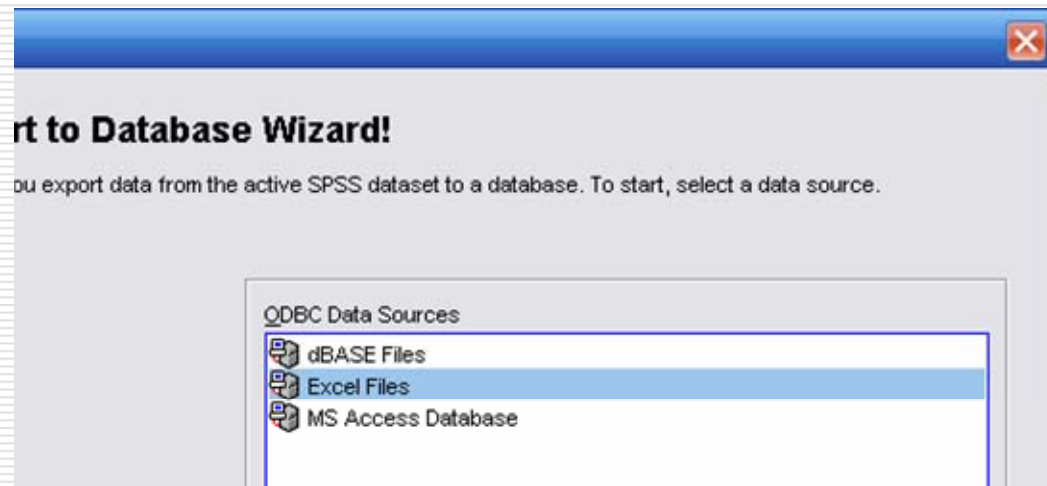
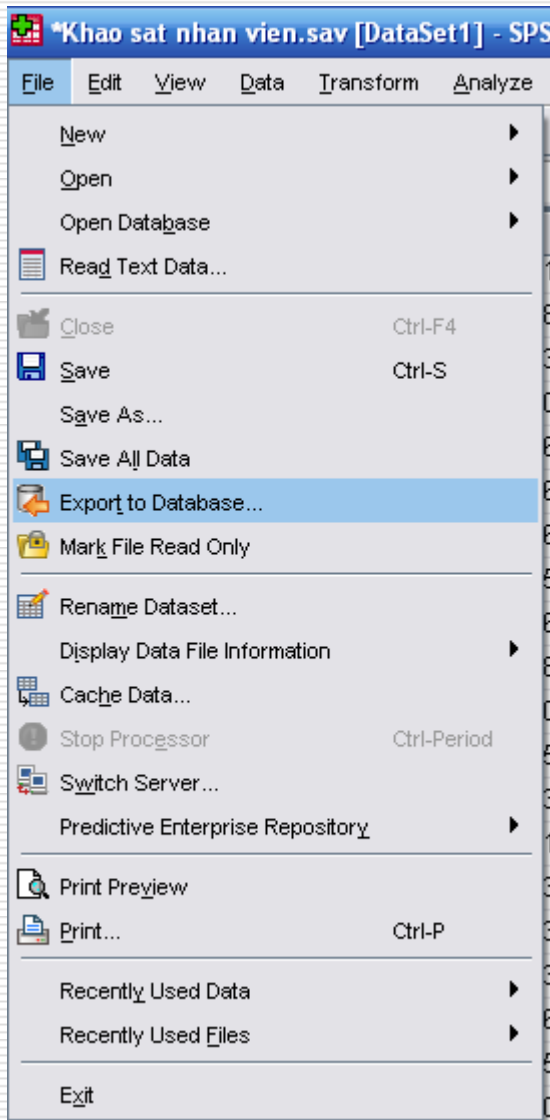
MÃ HÓA LẠI BIẾN



TÍNH TOÁN CÁC BIẾN



ĐỌC DỮ LIỆU TỪ EXCEL



PHÂN TÍCH MÔ TẢ

The screenshot shows the SPSS Data Editor interface. The menu bar includes File, Edit, View, Data, Transform, Analyze, Graphs, Utilities, Add-ons, Window, and Help. The 'Analyze' menu is open, and 'Frequencies...' is selected under 'Descriptive Statistics'. The background shows a data table with columns TUOI, TGLV, and THUNHAP.

	TUOI	TGLV	THUNHAP
150	26	1	
151	21	1	
152	30	1	
153	23	1	
154	26	1	
155	26	1	
156	22	1	
157	27	1	
158	26	1	

The screenshot shows the 'Frequencies' dialog box. The 'Variable(s):' list contains GIOITINH, CHMON, and TUOI. The 'Statistics...' and 'Charts...' buttons are highlighted with green arrows. The 'Display frequency tables' checkbox is checked.

Variable(s):
GIOITINH
CHMON
TUOI

Display frequency tables

OK Paste Reset Cancel Help

PHÂN TÍCH MÔ TẢ

Frequencies: Statistics

Percentile Values

- Quartiles
- Cut points for: 10 equal groups
- Percentile(s):
- Add
- Change
- Remove

Central Tendency

- Mean
- Median
- Mode
- Sum
- Values are group midpoints

Dispersion

- Std. deviation
- Minimum
- Variance
- Maximum
- Range
- S.E. mean

Distribution

- Skewness
- Kurtosis

Continue Cancel Help

Thống kê
xu hướng
hội tụ

Đồ thị

Frequencies: Charts

Chart Type

- None
- Bar charts
- Pie charts
- Histograms:
 - With normal curve

Chart Values

- Frequencies
- Percentages

Continue Cancel Help

Kiểm tra
độ phân
tán

Kiểm tra
phân phối

PHÂN TÍCH MÔ TẢ

GIOITINH

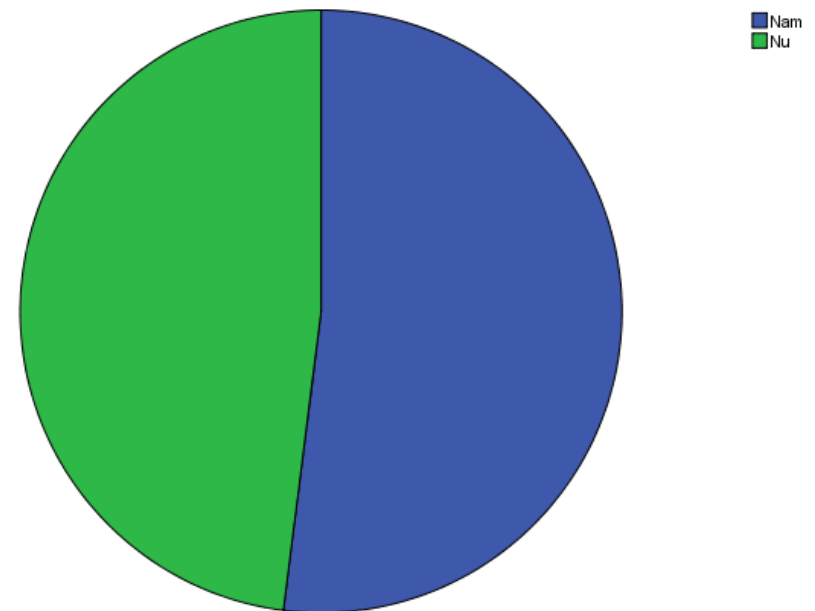
		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1	104	52.0	52.0	52.0
	2	96	48.0	48.0	100.0
	Total	200	100.0	100.0	

Statistics

TUOI

N	Valid	200
	Missing	0
Mean		26.47
Median		26.00
Minimum		21
Maximum		42
Sum		5294

Giới tính



NHẬP LIỆU CHO CÂU HỎI NHIỀU LỰA CHỌN

- Ví dụ:

- ⊕ Bạn hãy cho biết nhiều nhất 9 loại báo mà bạn thường đọc tại Cơ quan (nơi làm việc)

- ◆ Hà Nội mới
- ◆ SGGP
- ◆ Lao Động
- ◆ Tiền Phong
- ◆ Tuổi trẻ
- ◆ Phụ nữ Việt Nam
- ◆ Sài Gòn Tiếp thị
- ◆ Thế giới Phụ nữ
- ◆ An ninh thế giới

→ Có thể chọn nhiều hơn một phương án → làm thế nào để nhập liệu

NHẬP LIỆU CHO CÂU HỎI NHIỀU LỰA CHỌN

*THUC HANH SPSS - MBA.sav [DataSet1] - SPSS Data Editor

File Edit View Data Transform Analyze Graphs Utilities Add-ons Window Help

	Name	Type	Width	Decimals	Label
51	BAO1	Numeric	5	0	Báo thường đọc
52	BAO2	Numeric	6	0	Báo thường đọc
53	BAO3	Numeric	5	0	Báo thường đọc
54	BAO4	Numeric	5	0	Báo thường đọc
55	BAO5	Numeric	5	0	Báo thường đọc
56	BAO6	Numeric	5	0	Báo thường đọc
57	BAO7	Numeric	5	0	Báo thường đọc
58	BAO8	Numeric	5	0	Báo thường đọc
59	BAO9	Numeric	5	0	Báo thường đọc

Value Labels

Value:

Label:

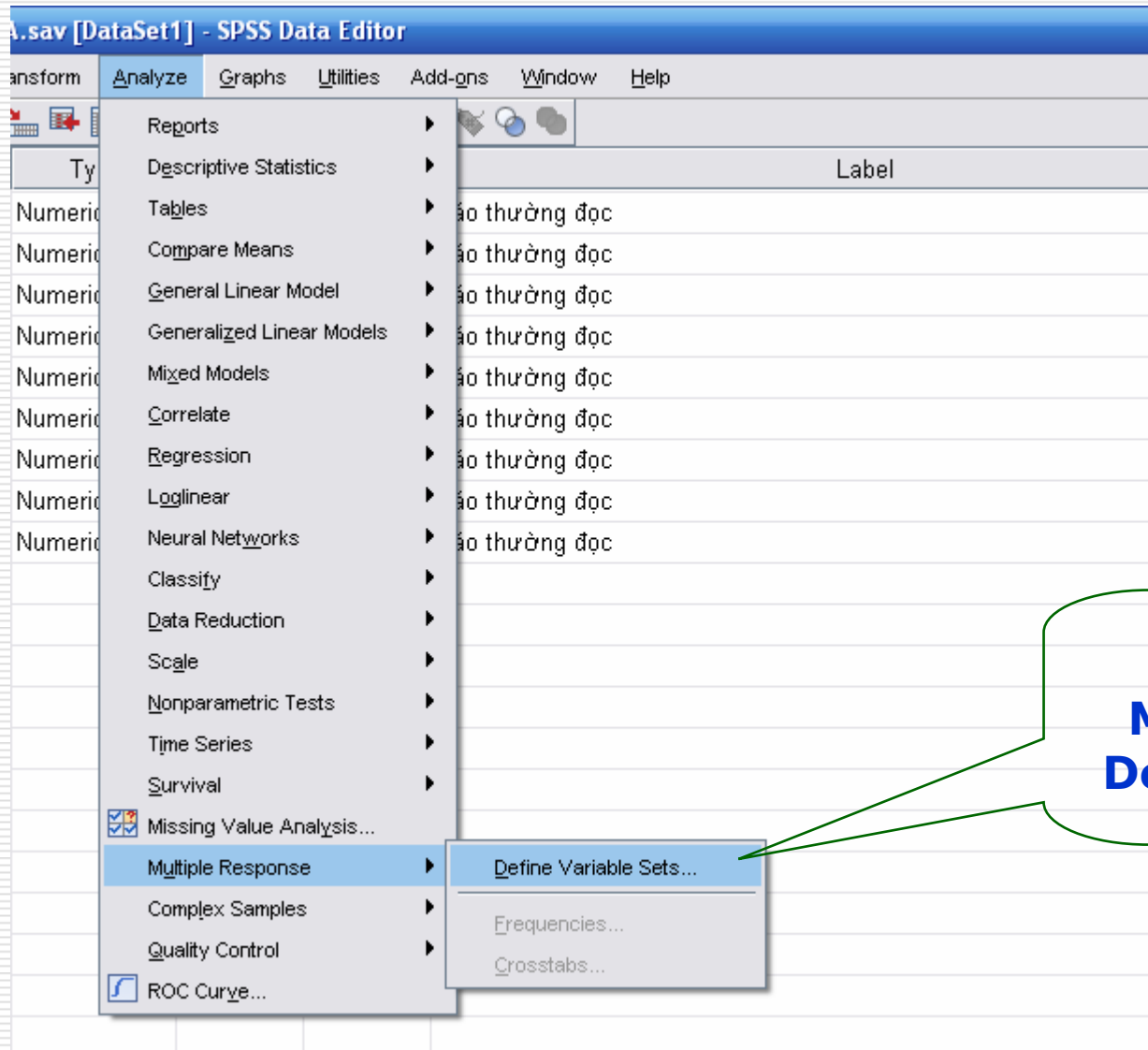
Spelling...

Add Change Remove

1 = "Hà Nội mới"
2 = "SGGP"
3 = "Lao Động"
4 = "Người Lao động"
5 = "Tiền Phong"
6 = "Thanh Niên"
7 = "Tuổi trẻ"

OK Cancel Help

THỐNG KÊ CHO CÂU HỎI NHIỀU LỰA CHỌN



**Analyze
Multiple Response
Define Variable Sets**

NHẬP LIỆU CHO CÂU HỎI NHIỀU LỰA CHỌN

Set Definition

Thu nhập [THUNHAP]
Lãnh đạo [F1]
Cơ hội đào tạo và th...
Lương và thưởng [F3]
Đồng nghiệp [F4]
Phúc lợi [F5]
Bản chất công việc [...]
Môi trường làm việc [...]
Lòng trung thành [LTT]

Variables in Set:

Báo thường đọc [BA...]
Báo thường đọc [BA...]
Báo thường đọc [BA...]
Báo thường đọc [BA...]
Báo thường đọc [BA...]
Báo thường đọc [BA...]
Báo thường đọc [BA...]
Báo thường đọc [BA...]
Báo thường đọc [BA...]
Báo thường đọc [BA...]

Variables Are Coded As

Dichotomies Counted value:

Categories Range: 1 through 18

Name: BAO
Label: Báo thường đọc

Multiple Response Sets:

\$BAO

Add
Change
Remove

Note: Sets defined here are only available in the Multiple Response Frequencies and Crosstabs procedures. Use Define Multiple Response Sets on the Data menu for sets used elsewhere.

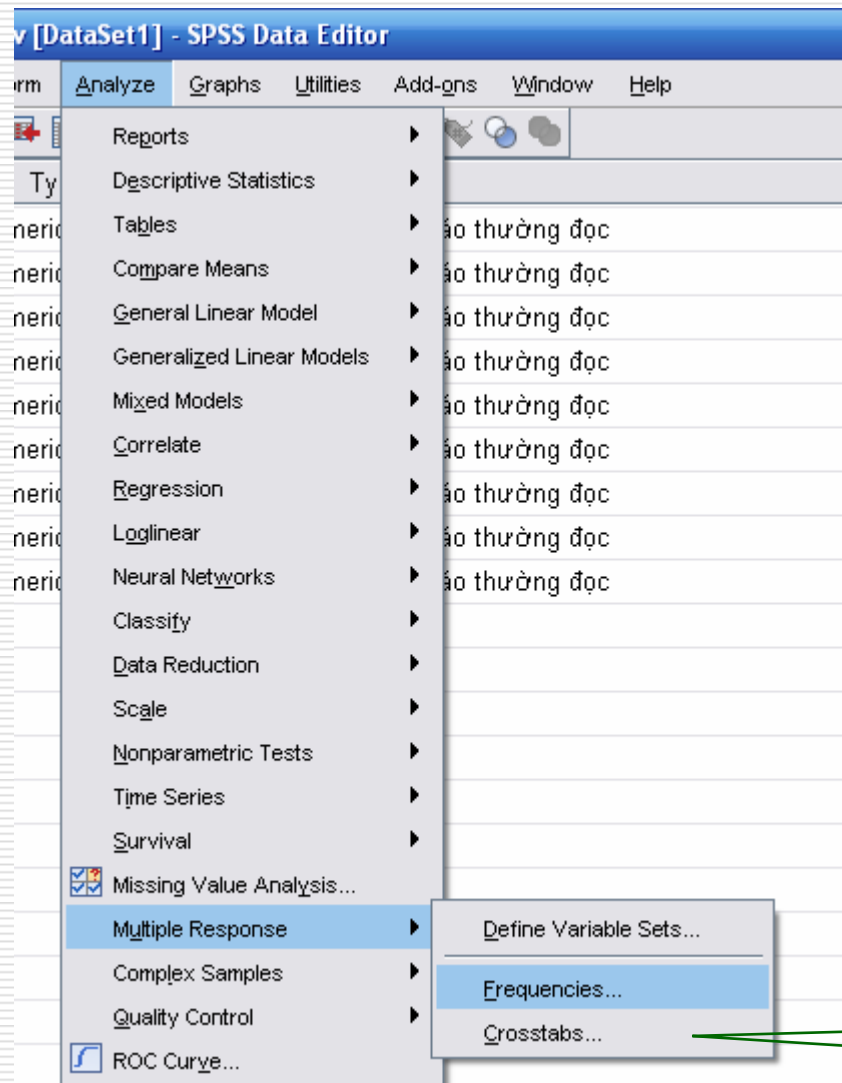
Close Help

Nhấn Add

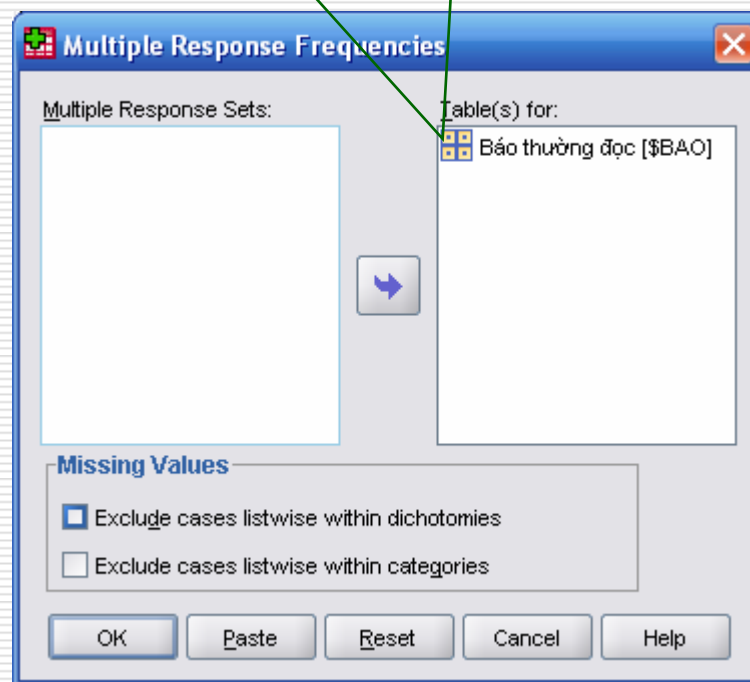
**Nhập kiểu định nghĩa
Dichotomies / Categories
Nhập số lượng lựa chọn**

**Nhập thông tin
Biến: BAO
Label: Báo thường
đọc**

THỐNG KÊ CHO CÂU HỎI NHIỀU LỰA CHỌN



Chọn và chuyển tên
biến cần tính



Mô tả tần suất

THỐNG KÊ CHO CÂU HỎI NHIỀU LỰA CHỌN

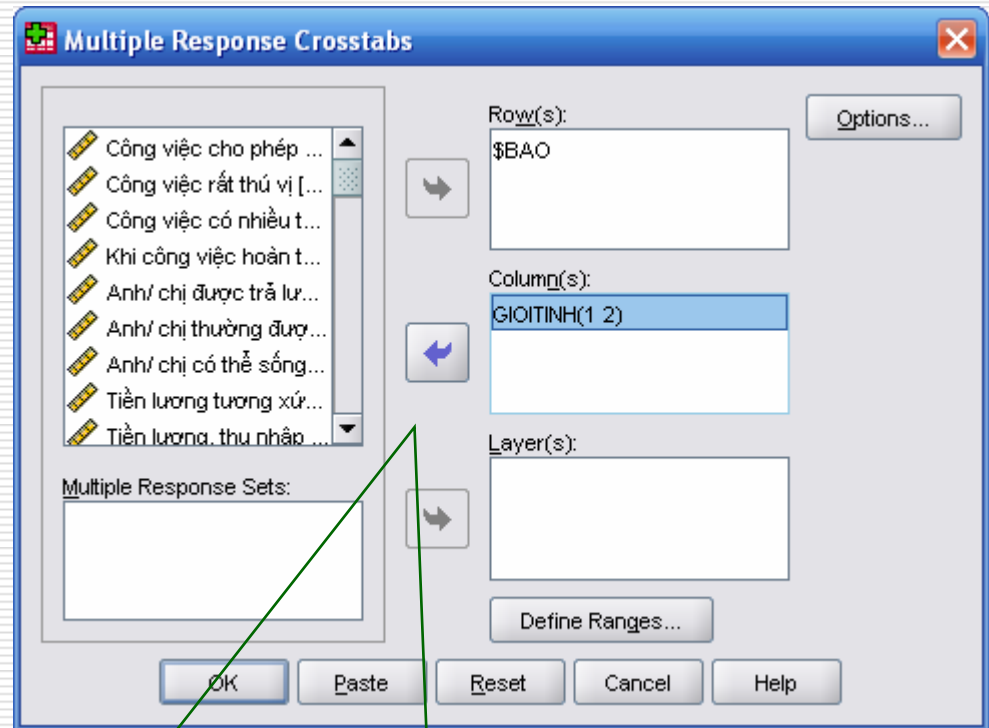
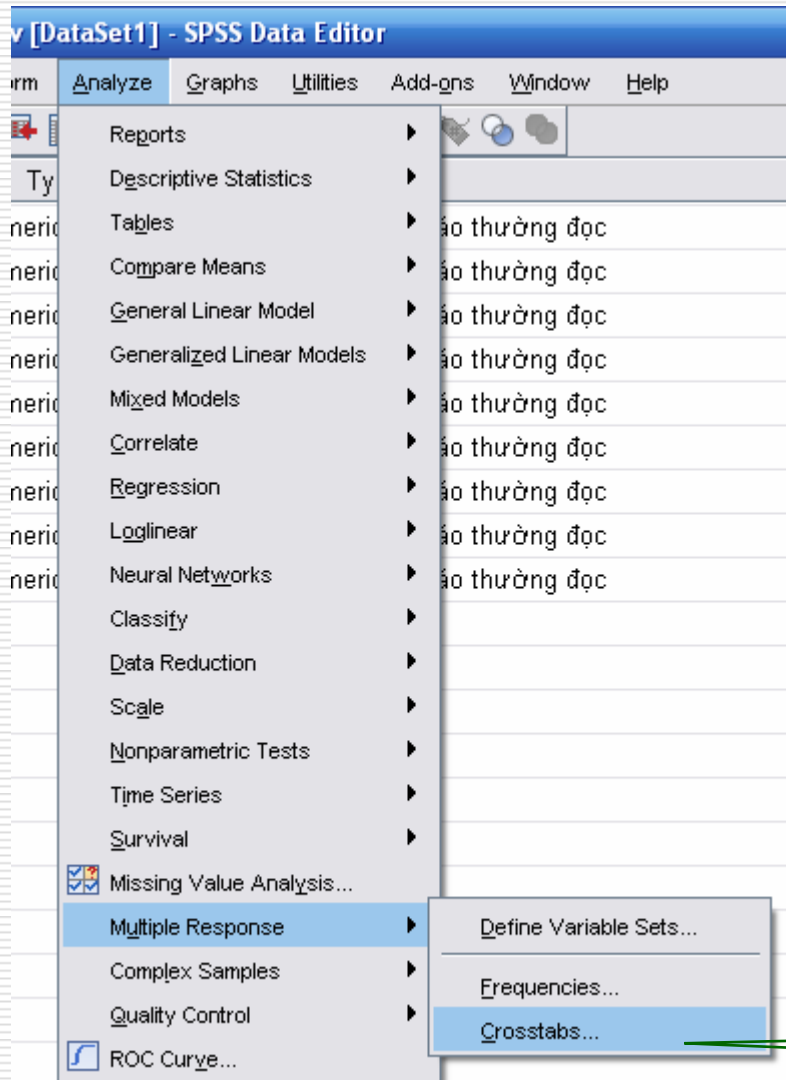
\$BAO Frequencies

		Responses		Percent of Cases
		N	Percent	
Báo thường đọc ^a	Hà Nội mới	134	15.3%	67.0%
	SGGP	6	.7%	3.0%
	Lao Động	54	6.2%	27.0%
	Người Lao động	18	2.1%	9.0%
	Tiền Phong	36	4.1%	18.0%
	Thanh Niên	24	2.7%	12.0%
	Tuổi trẻ	36	4.1%	18.0%
	Phụ nữ Việt Nam	80	9.1%	40.0%
	Thời báo KTSG	30	3.4%	15.0%
	Thời báo KT Việt Nam	6	.7%	3.0%
	Sài Gòn Tiếp thị	42	4.8%	21.0%
	Thế giới Phụ nữ	48	5.5%	24.0%
	Tiếp thị và cuộc sống	18	2.1%	9.0%
	Mua và bán	12	1.4%	6.0%
	An ninh thế giới	156	17.8%	78.0%
	An ninh thủ đô	170	19.4%	85.0%
	Công An TP HCM	6	.7%	3.0%
	Total		876	100.0%

Kết quả thống kê

a. Group

THỐNG KÊ CHO CÂU HỎI NHIỀU LỰA CHỌN



**Chọn các biến vào phân tích
Nhân Define Ranges để định
nghĩa biến nominal ở cột**

Kết quả thống kê

THỐNG KÊ CHO CÂU HỎI NHIỀU LỰA CHỌN

\$BAO*GIOITINH Crosstabulation

			Giới tính		Total
			Nam	Nu	
Báo thường đọc ^a	Hà Nội mới	Count	71	63	134
	SGGP	Count	1	5	6
	Lao Động	Count	33	21	54
	Người Lao động	Count	3	15	18
	Tiền Phong	Count	20	16	36
	Thanh Niên	Count	12	12	24
	Tuổi trẻ	Count	21	15	36
	Phụ nữ Việt Nam	Count	37	43	80
	Thời báo KTSG	Count	15	15	30
	Thời báo KT Việt Nam	Count	1	5	6
	Sài Gòn Tiếp thị	Count	18	24	42
	Thế giới Phụ nữ	Count	26	22	48
	Tiếp thị và cuộc sống	Count	12	6	18
	Mua và bán	Count	6	6	12
	An ninh thế giới	Count	79	77	156
	An ninh thủ đô	Count	88	82	170
	Công An TP HCM	Count	3	3	6
Total	Count	104	96	200	

Percentages and totals are based on respondents.

a. Group

HÌNH THÀNH CÁC BẢNG BIỂU (Custom Tables)

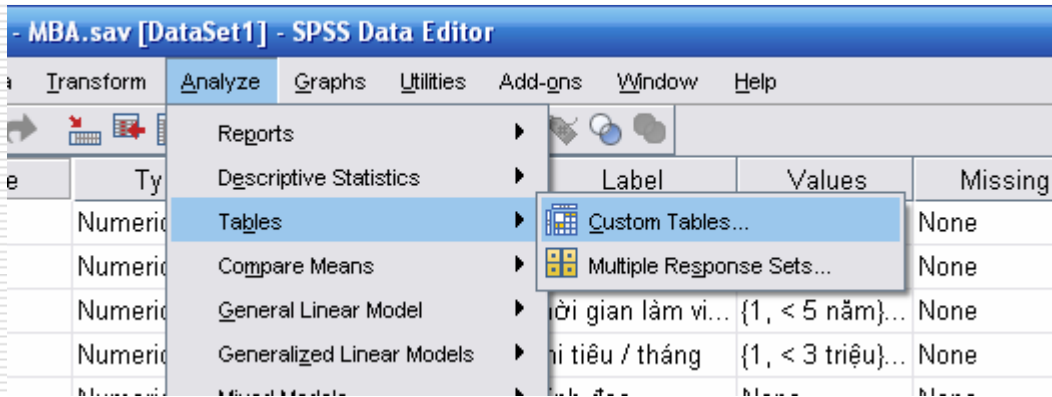
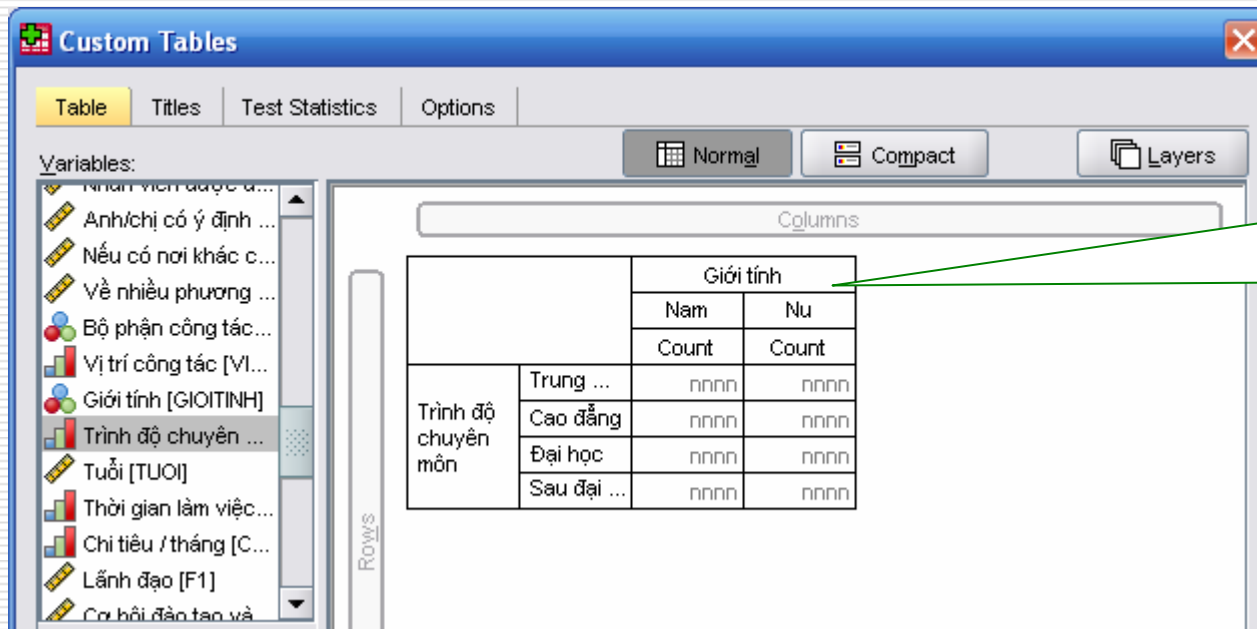


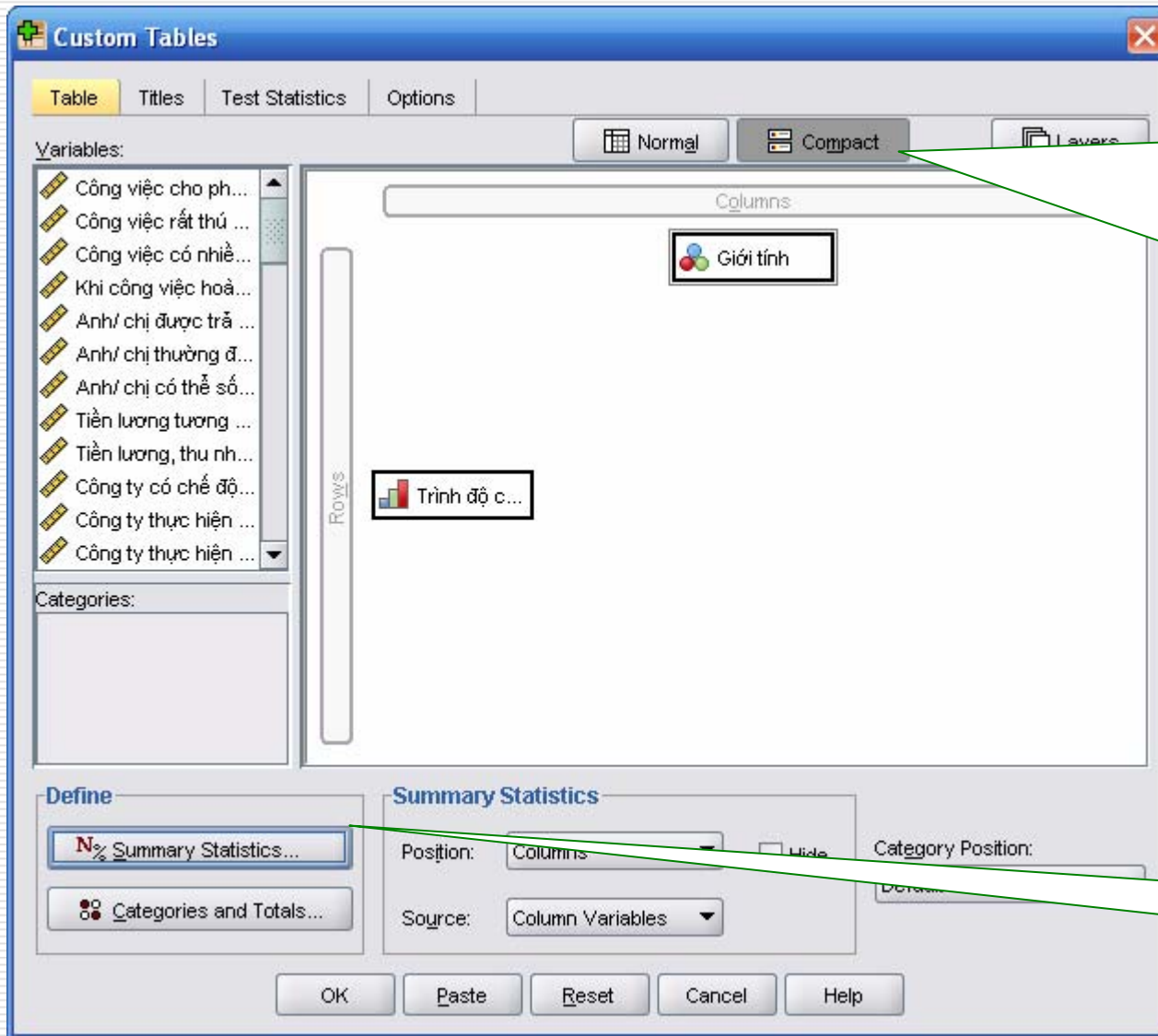
Table 1

		Giới tính	
		Nam	Nu
Trình độ chuyên môn	Trung cấp	11	3
	Cao đẳng	3	8
Đại học	87	82	
Sau đại học	3	3	



**Kéo biến GIOITINH
sang ô cột, biến
TRINHDO sang ô
dòng**

HÌNH THÀNH CÁC BẢNG BIỂU (Custom Tables)

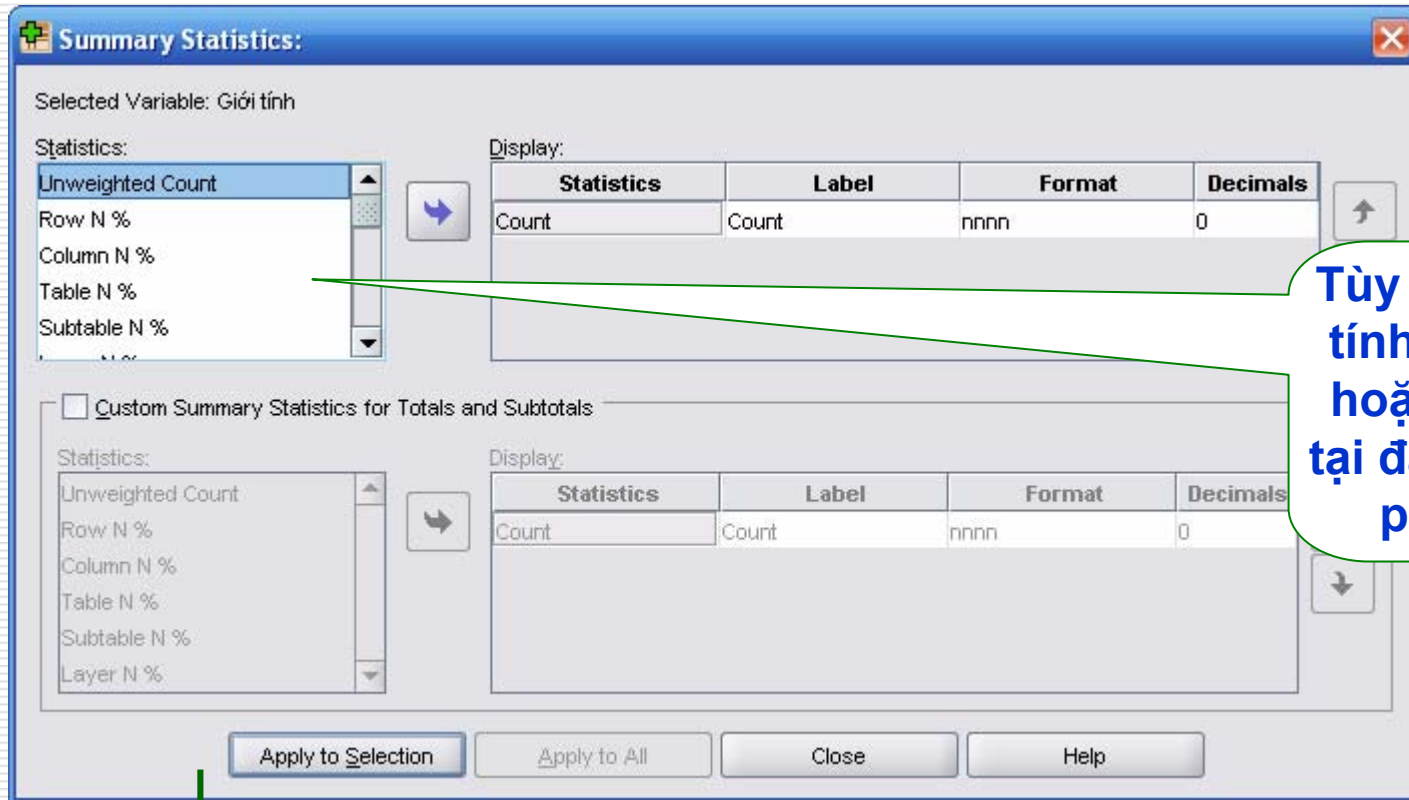


**Kéo biến GIOITINH
sang ô cột, biến
TRINHDO sang ô
dòng**

**Nếu tính thêm %
thì chọn Compact**

**Sau đó chọn
Summary Statistics**

HÌNH THÀNH CÁC BẢNG BIỂU (Custom Tables)



Tùy theo mục đích tính % theo hàng, hoặc cột thì chọn tại đây và kéo sang phần Display

Table 1

		Giới tính					
		Nam			Nu		
		Count	Row N %	Column N %	Count	Row N %	Column N %
Trình độ chuyên môn	Trung cấp	11	78.6%	10.6%	3	21.4%	3.1%
	Cao đẳng	3	27.3%	2.9%	8	72.7%	8.3%
	Đại học	87	51.5%	83.7%	82	48.5%	85.4%
	Sau đại học	3	50.0%	2.9%	3	50.0%	3.1%

Sau đó chọn Apply to Selection

HÌNH THÀNH CÁC BẢNG BIỂU (Custom Tables)

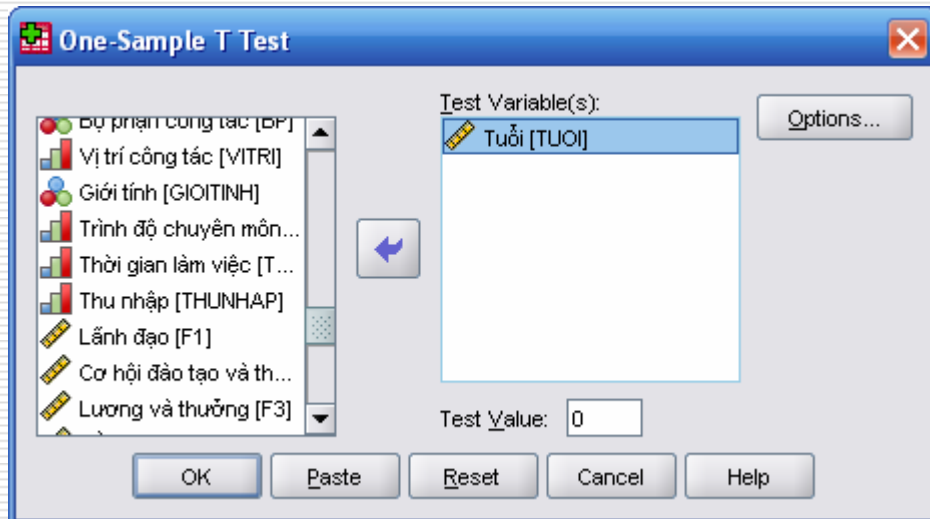
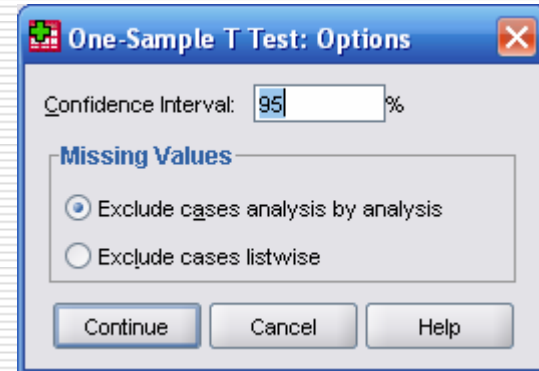
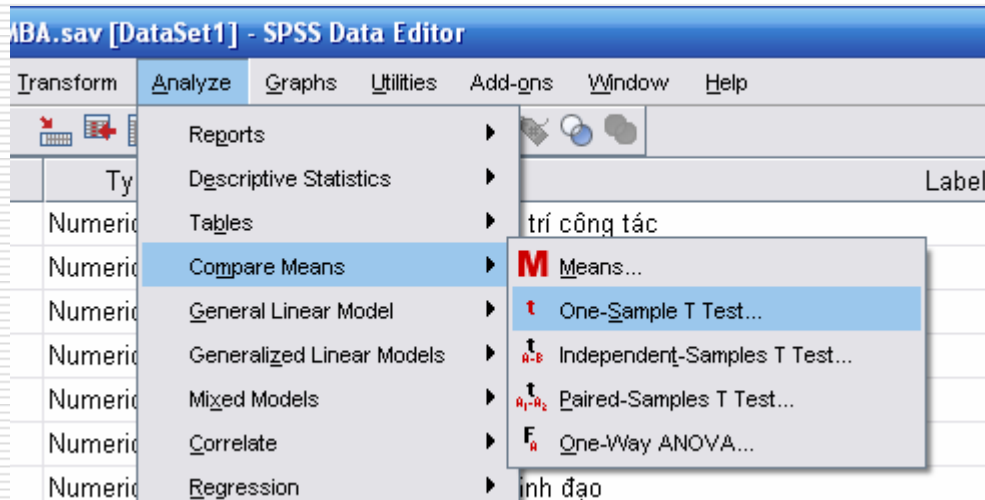
Table 1

		Giới tính					
		Nam			Nu		
		Count	Row N %	Column N %	Count	Row N %	Column N %
Trình độ chuyên môn	Trung cấp	11	78.6%	10.6%	3	21.4%	3.1%
	Cao đẳng	3	27.3%	2.9%	8	72.7%	8.3%
	Đại học	87	51.5%	83.7%	82	48.5%	85.4%
	Sau đại học	3	50.0%	2.9%	3	50.0%	3.1%

ƯỚC LƯỢNG THAM SỐ TRUNG BÌNH

ƯỚC LƯỢNG KHOẢNG TIN CẬY CỦA GIÁ TRỊ TRUNG BÌNH	
Đã biết phương sai tổng thể (σ^2)	Chưa biết phương sai tổng thể (σ^2)
<p><i>Phân phối chuẩn</i> σ: Độ lệch chuẩn của tổng thể μ: Trung bình tổng thể $\alpha = 1 - \gamma$: Mức ý nghĩa</p> $\bar{x} - \frac{\sigma}{\sqrt{n}} U_{1-\frac{\alpha}{2}} < \mu < \bar{x} + \frac{\sigma}{\sqrt{n}} U_{1-\frac{\alpha}{2}}$	<p><i>Phân phối Student với (n-1) là bậc tự do</i> s': Độ lệch chuẩn hiệu chỉnh mẫu</p> $\bar{x} - \frac{s'}{\sqrt{n}} t_{\frac{\alpha}{2}}^{n-1} < \mu < \bar{x} + \frac{s'}{\sqrt{n}} t_{\frac{\alpha}{2}}^{n-1}$
	<p><i>Phân phối chuẩn</i> s': Độ lệch chuẩn hiệu chỉnh mẫu</p> $\bar{x} - \frac{s'}{\sqrt{n}} U_{1-\frac{\alpha}{2}} < \mu < \bar{x} + \frac{s'}{\sqrt{n}} U_{1-\frac{\alpha}{2}}$

ƯỚC LƯỢNG THAM SỐ TRUNG BÌNH



ƯỚC LƯỢNG THAM SỐ TRUNG BÌNH

One-Sample Statistics

	N	Mean	Std. Deviation	Std. Error Mean
Tuổi	200	26.47	3.161	.223

One-Sample Test

	Test Value = 0					
	t	df	Sig. (2-tailed)	Mean Difference	95% Confidence Interval of the Difference	
					Lower	Upper
Tuổi	118.442	199	.000	26.470	26.03	26.91

SO SÁNH 2 THAM SỐ TRUNG BÌNH

- Ước lượng tham số trung bình:
 - ⊕ Dữ liệu: định lượng (liên tục)
 - ⊕ Độ tin cậy (1-mức ý nghĩa)
- So sánh trung bình hai mẫu độc lập:
 - ⊕ H_0 : Không khác nhau giữa độ tuổi trung bình giữa người nữ và người nam trong tổng thể
 - H_1 : Có sự khác nhau giữa độ tuổi trung bình giữa người nữ và người nam trong tổng thể
 - ⊕ Điều kiện ứng dụng:
 - ◆ Dữ liệu định lượng (liên tục)
 - ◆ Dữ liệu của hai nhóm phải tuân theo quy luật phân phối chuẩn

SO SÁNH 2 THAM SỐ TRUNG BÌNH

KIỂM ĐỊNH THAM SỰ KHÁC NHAU HAI TRUNG BÌNH TỔNG THỂ

(dựa trên sự phân phối từng cặp)

1. Giả thiết và đối thiết:

	Đôi xứng	Phải	Trái
Giả thiết	$H_0: \mu_x - \mu_y = D_0$	$H_0: \mu_x - \mu_y \leq D_0$	$H_0: \mu_x - \mu_y \geq D_0$
Đối thiết	$H_1: \mu_x - \mu_y \neq D_0$	$H_1: \mu_x - \mu_y > D_0$	$H_1: \mu_x - \mu_y < D_0$

2. Xác định mức ý nghĩa




3. Phương pháp kiểm nghiệm sự khác nhau của hai trung bình tổng thể - Bảng phân phối chuẩn hoặc T-student (nếu $n < 30$)

4. Tiêu chuẩn kiểm định T hoặc U:

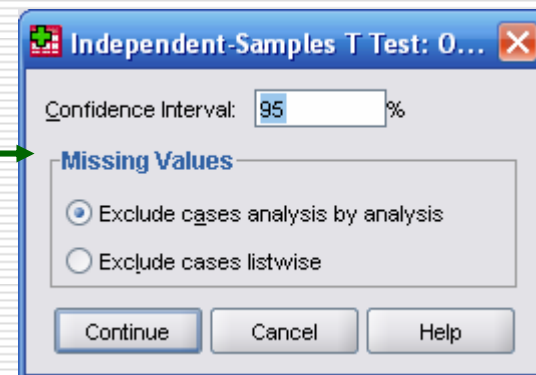
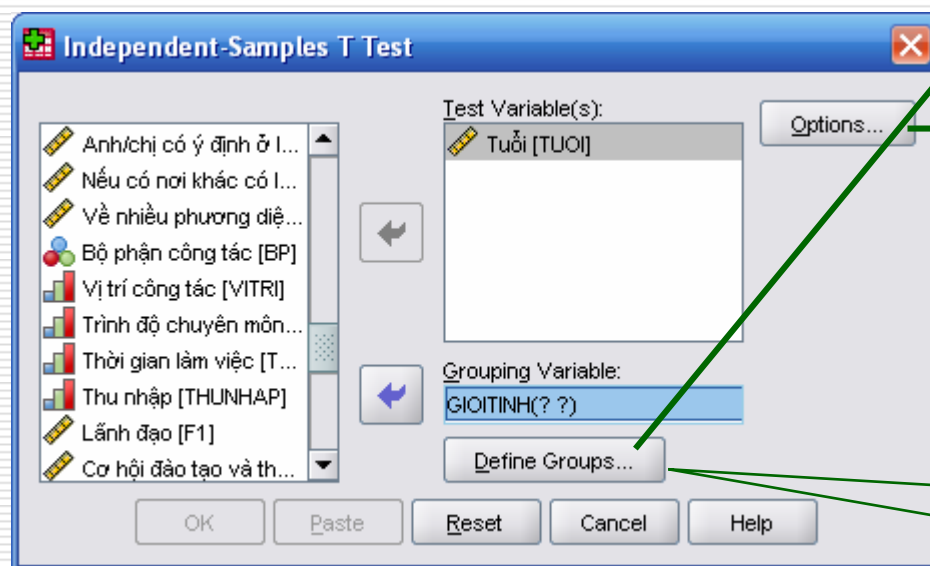
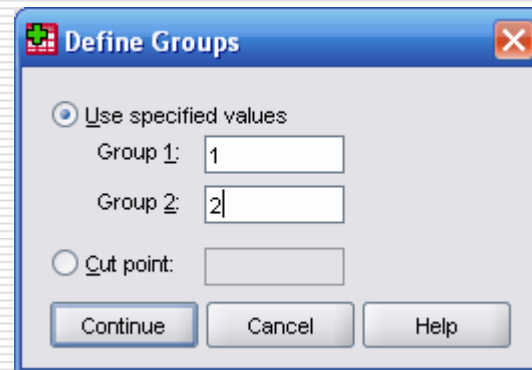
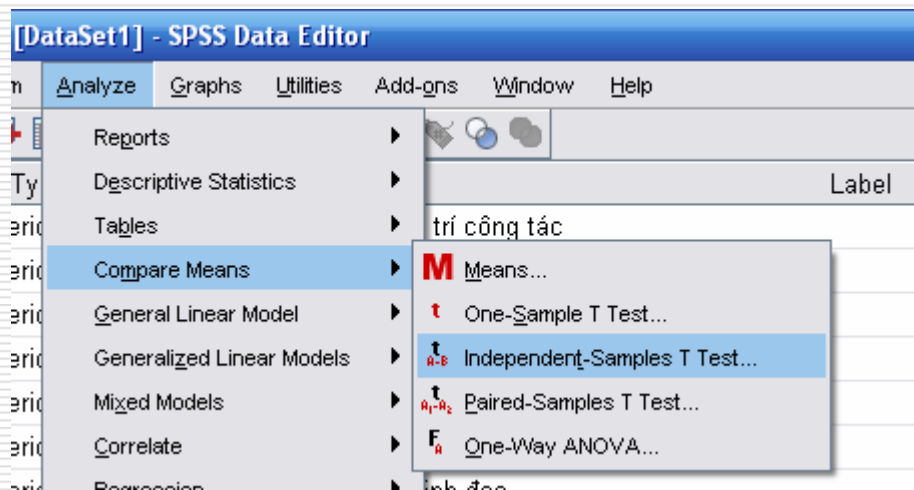
$$k_{qr} \equiv T = \frac{(\bar{x} - D_0) \sqrt{n}}{s'_d}$$

\bar{x} và s'_d là trung bình và độ lệch chuẩn điều chỉnh của n khác biệt

5. Điểm tới hạn và miền bác bỏ:

	Đôi xứng	Phải	Trái
Điểm tới hạn	$-T_{(n-1); 1-\alpha/2}$ và $T_{(n-1); 1-\alpha/2}$	$T_{(n-1); 1-\alpha}$	$-T_{(n-1); 1-\alpha}$
Miền bác bỏ	$D < -T_{(n-1); 1-\alpha/2}$ và $D > T_{(n-1); 1-\alpha/2}$	$D > T_{(n-1); 1-\alpha}$	$D < -T_{(n-1); 1-\alpha}$
Mô hình			

TRÌNH TỰ THỰC HIỆN



**Chọn các biến phân tích
Nhấn Define Group để Đ.ng**

KẾT QUẢ PHÂN TÍCH

Group Statistics

	Giới tính	N	Mean	Std. Deviation	Std. Error Mean
Tuổi	Nam	104	27.06	3.353	.329
	Nu	96	25.83	2.820	.288

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means					95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
Tuổi	Equal variances assumed	.345	.558	2.783	198	.006	1.224	.440	.357	2.092
	Equal variances not assumed			2.802	196.340	.006	1.224	.437	.363	2.086

Nếu sig. trong kiểm định phương sai > 0,05 thì phương sai giữa hai mẫu bằng nhau, ta sẽ dùng kết quả kiểm định t ở dòng thứ 1

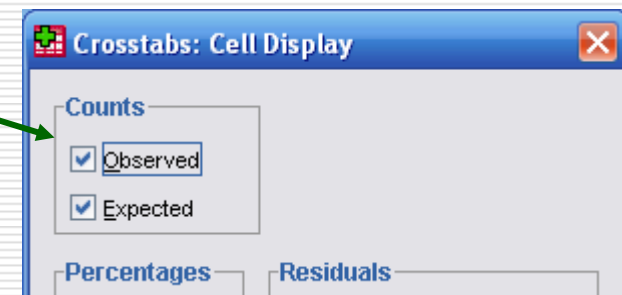
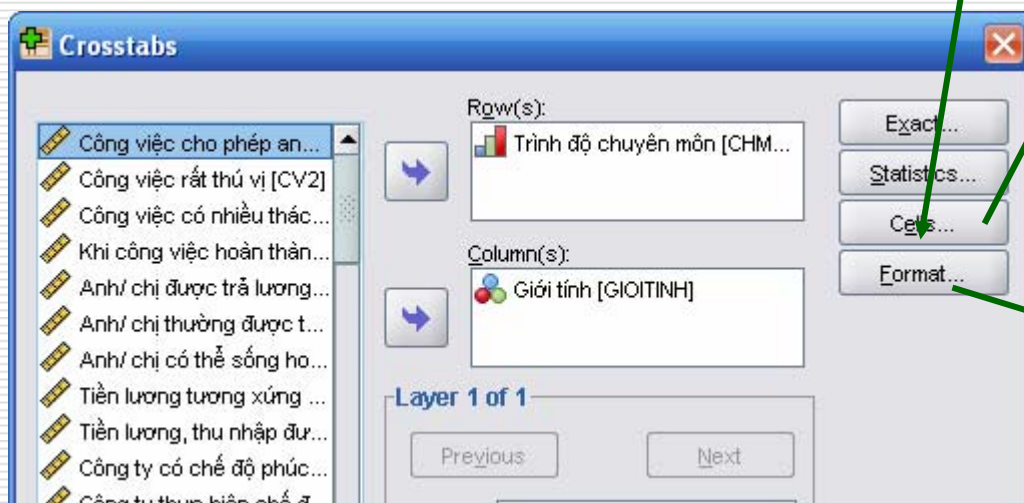
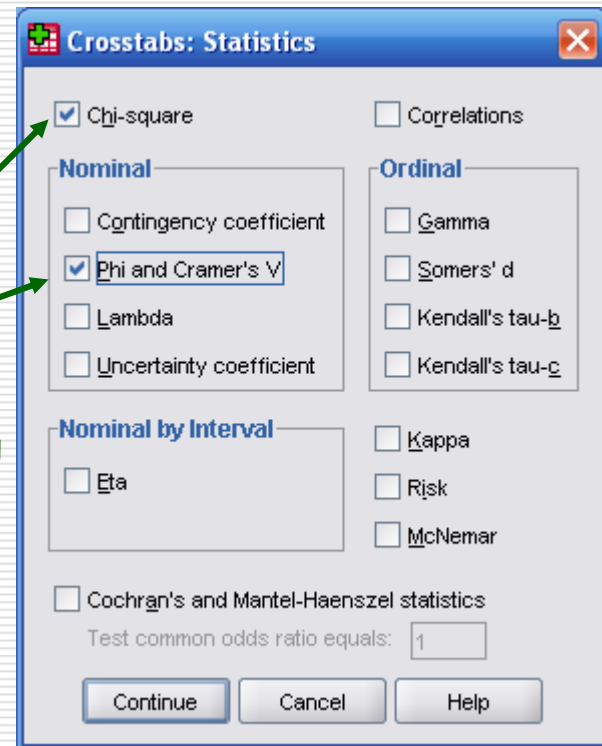
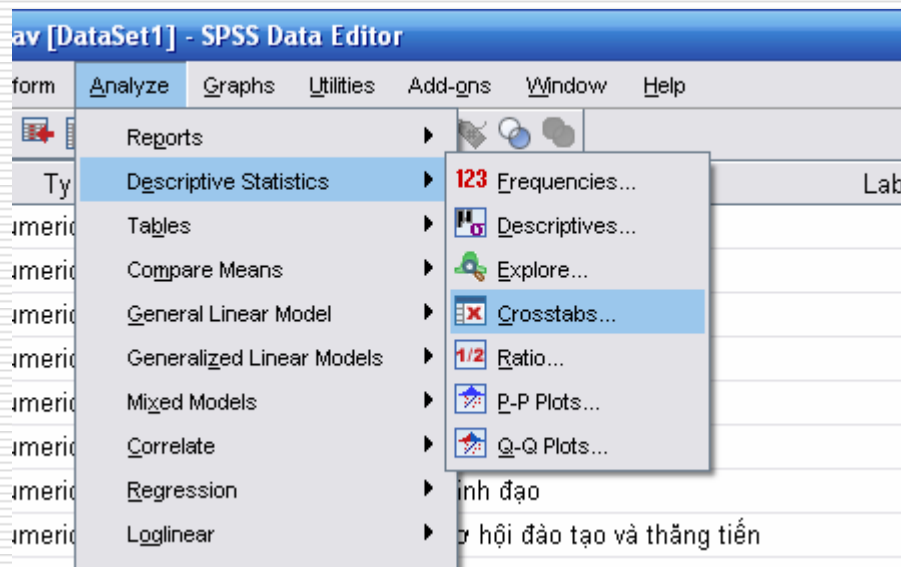
Giá trị t của kiểm định

p-value của giá trị t

PHÂN TÍCH MQH GIỮA 2 BIẾN ĐỊNH TÍNH

- Điều kiện (hai biến định tính)
- Các giả thuyết:
 - H_0 : Hai biến độc lập với nhau (hai biến không có MQH)*
 - H_1 : Hai biến có liên quan với nhau*
- χ^2 được thiết lập để xác định có hay không một mối liên hệ giữa hai biến, nhưng nó không chỉ ra được cường độ của mối liên hệ đó.
- χ^2 cho phép tìm ra những mối liên hệ phi tuyến tính
- Cramer-V: Cường độ của nó biến động từ 0 đến 1.

PHÂN TÍCH MQH GIỮA 2 BIẾN ĐỊNH TÍNH



PHÂN TÍCH MQH GIỮA 2 BIẾN ĐỊNH TÍNH

Symmetric Measures

		Value	Approx. Sig.
Nominal by Nominal	Phi	.183	.083
	Cramer's V	.183	.083
N of Valid Cases		200	

Trình độ chuyên môn * Giới tính Crosstabulation

			Giới tính		Total
			Nam	Nu	
Trình độ chuyên môn	Trung cấp	Count	11	3	14
		Expected Count	7.3	6.7	14.0
	Cao đẳng	Count	3	8	11
		Expected Count	5.7	5.3	11.0
	Đại học	Count	87	82	169
		Expected Count	87.9	81.1	169.0
	Sau đại học	Count	3	3	6
		Expected Count	3.1	2.9	6.0
Total		Count	104	96	200
		Expected Count	104.0	96.0	200.0

CHƯƠNG 2

HƯỚNG DẪN SỬ DỤNG EIEWS TRONG PHÂN TÍCH DỮ LIỆU VÀ HỒI QUI

Chương này sẽ trình bày một số thủ tục cơ bản của phần mềm Eviews 5.1 để sinh viên có thể thực hành các bài tập thống kê và kinh tế lượng ở các chương sau. Do mục đích chính của ta là thực hành kinh tế lượng với Eviews, nên chương này chỉ giới hạn một số thao tác mà người nghiên cứu thường hay sử dụng, chứ không phải toàn bộ hướng dẫn chi tiết cách sử dụng Eviews. Tuy nhiên, để tiện lợi cho sinh viên tự nghiên cứu, chương này sẽ giới thiệu sơ qua chức năng trợ giúp trong Eviews để có thể tham khảo khi cần thiết. Một số nội dung được trình bày trong chương này, đặc biệt là các kiểm định, nhưng chúng sẽ được hướng dẫn một cách chi tiết hơn ở các chương liên quan. Để sinh viên có thể thực hành các bài tập và dự án nghiên cứu với Eviews, chương này sẽ nhằm vào các nội dung sau đây:

- Eviews là gì?
- Workfile là gì?
- Trình bày dữ liệu trong Eviews
- Đối tượng trong Eviews
- Quản lý dữ liệu trong Eviews
- Các phép toán và hàm số trong Eviews?
- Phân tích dữ liệu chuỗi và nhóm
- Xây dựng hàm kinh tế lượng trong Eviews
- Kiểm định giả thiết mô hình hồi qui trong Eviews

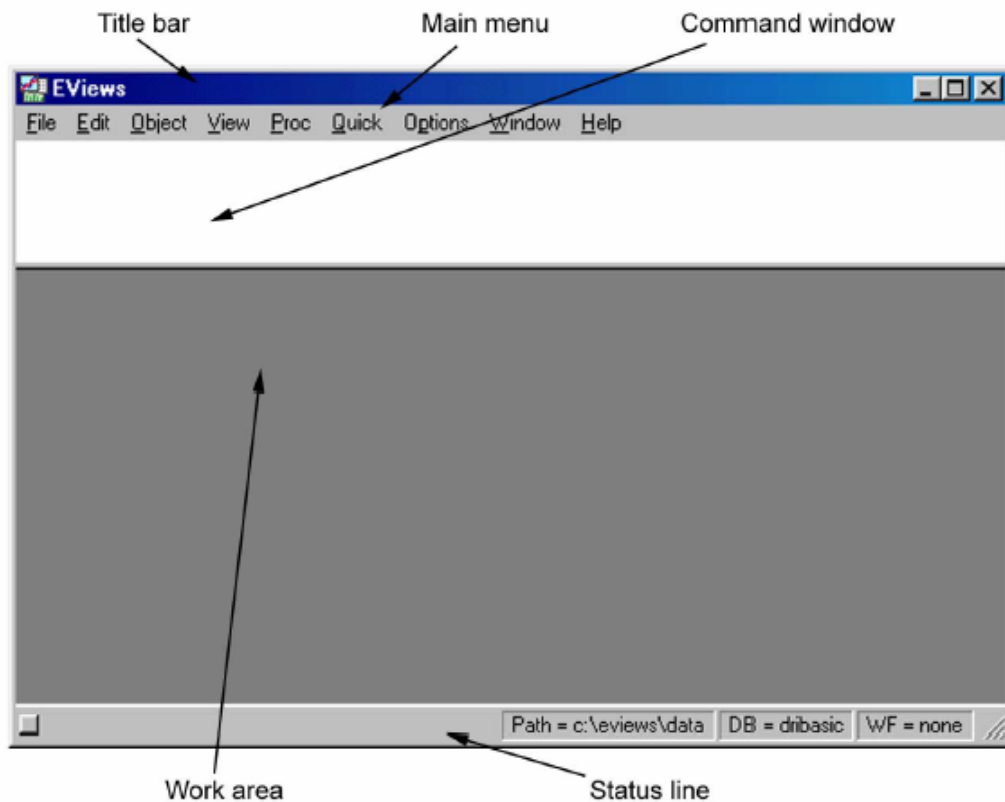
NHỮNG VẤN ĐỀ CƠ BẢN VỀ EIEWS

EIEWS LÀ GÌ?

Eviews¹ cung cấp các công cụ phân tích dữ liệu phức tạp, hồi qui và dự báo chạy trên Windows. Với Eviews ta có thể nhanh chóng xây dựng một mối quan hệ kinh tế lượng từ dữ liệu có sẵn và sử dụng mối quan hệ này để dự báo các giá trị tương lai. Eviews có thể hữu ích trong tất cả các loại nghiên cứu như đánh giá và phân tích dữ liệu khoa học, phân tích tài chính, mô phỏng và dự báo vĩ mô, dự báo doanh số, và phân tích chi phí. Đặc biệt, Eviews là một phần mềm rất mạnh cho phân tích dữ liệu thời gian cũng như chéo với cỡ mẫu lớn.

¹ Viết tắt của Econometrics Views

Eviews đưa ra nhiều cách nhập dữ liệu rất thông dụng và dễ sử dụng như nhập từ bàn phím, từ các file sẵn có dưới dạng excel hay text, dễ dàng mở rộng file dữ liệu có sẵn. Eviews trình bày các biểu mẫu, đồ thị, kết quả ấn tượng và có thể in trực tiếp hoặc chuyển qua các loại định dạng văn bản khác. Eviews giúp người sử dụng dễ dàng ước lượng và kiểm định các mô hình kinh tế lượng. Ngoài ra, Eviews còn giúp người nghiên cứu có thể xây dựng các file chương trình cho dự án nghiên cứu của mình. Eviews tận dụng các đặc điểm hiển thị của phần mềm Windows hiện đại nên rất thuận tiện cho người sử dụng như dùng chuột, các thanh kéo, thay đổi giao diện, thoát, ... Nếu chương trình được cài đặt đúng, thì khi khởi động evIEWS ta sẽ thấy cửa sổ chính như sau:



Nguồn: Eviews 5 Users Guide, pp.16

WORKFILE LÀ GÌ?

Workfile được gọi chung là tập tin làm việc của Eviews (sau đây sẽ gọi là tập tin Eviews). Ở một cấp độ cơ bản, một tập tin Eviews đơn giản là một tập tin chứa các đối tượng của Eviews¹. Mỗi đối tượng bao gồm một tập hợp các thông tin có liên quan đến một lĩnh vực phân tích cụ thể ví dụ một chuỗi², một phương trình, hay một đồ thị. Làm việc trên Eviews chủ yếu liên quan đến các đối tượng chứa trong một tập tin Eviews. Cho nên trước hết cần tạo một tập tin mới hoặc mở một tập tin có sẵn. Mỗi tập tin Eviews chứa một hoặc nhiều trang³. Mỗi trang chứa các đối tượng riêng. Trang

¹ Container for Eviews objects

² Series

³ Page

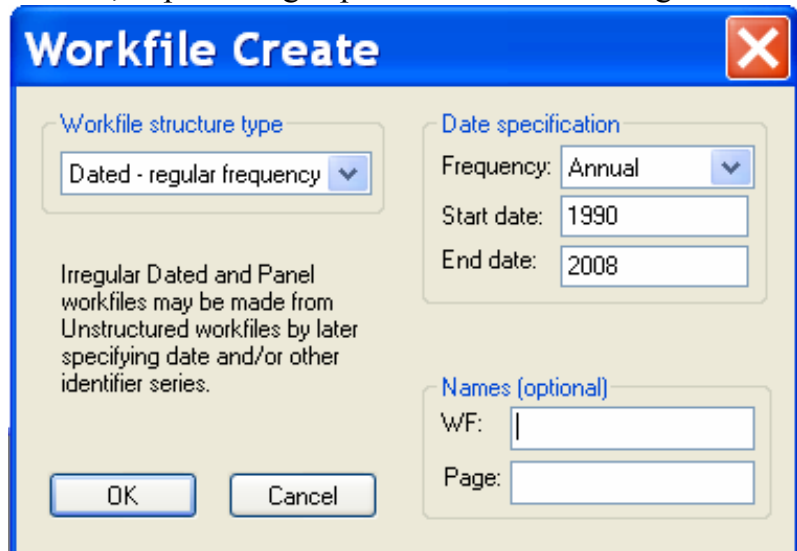
được xem như một thư mục con hay tập tin phụ¹ trong một tập tin. Lưu ý, một tập hợp các quan sát của một hoặc một số biến được gọi là bộ dữ liệu, trong đó mỗi quan sát có nhận dạng riêng.

Tạo một tập tin Eviews

Có nhiều cách tạo một tập tin mới. Việc đầu tiên trong tạo tập tin là xác định cấu trúc của tập tin². Có ba cách khác nhau. Thứ nhất là mô tả cấu trúc của tập tin Eviews. Theo cách này, Eviews sẽ tạo ra một tập tin mới để ta nhập dữ liệu vào một cách thủ công từ bàn phím hoặc copy và dán. Thứ hai là mở và đọc dữ liệu từ một nguồn bên ngoài (không thuộc định dạng Eviews) như Text, Excel, Stata. Thuận tiện hơn nhiều so với các phiên bản trước, Eviews 5 tự động phân tích nguồn dữ liệu, tạo một tập tin, và nhập dữ liệu. Đây là cách được sử dụng phổ biến. Thứ ba là tạo một tập tin theo hai bước riêng biệt. Trong bước một ta tạo ra một tập tin mới theo một trong hai cách trên. Trong bước hai ta sẽ cấu trúc tập tin. Chương này chỉ tập trung hướng dẫn cách thứ nhất và thứ hai.

(i) Tạo một tập tin bằng cách mô tả cấu trúc

Để mô tả cấu trúc của tập tin Eviews, ta phải cung cấp cho Eviews các thông tin về số quan sát và các nhận dạng liên quan. Để tạo một tập tin mới trên Eviews, ta chọn **File/New Workfile**, ... từ thực đơn chính³ để mở hộp thoại **Workfile Create**. Ở góc trái của hộp thoại là một hộp nhỏ để mô tả cấu trúc cơ bản của bộ dữ liệu. Ta có thể chọn giữa **Dated-Regular Frequency**, **Unstructured**, và **Balanced Panel**. Nói chung, ta có thể sử dụng **Dated-regular frequency**⁴ nếu ta có bộ dữ liệu thời gian, với bộ dữ liệu bảng đơn giản ta sử dụng **Balanced Panel**, và các trường hợp khác ta sử dụng **Unstructured**⁵.



Sau khi ta đã xác định loại cấu trúc dữ liệu, Eviews sẽ tự động nhắc ta mô tả đặc điểm của bộ dữ liệu đó như tần suất, ngày bắt đầu, ngày kết thúc đối với loại dữ

¹ Subdirectory/Subworkfile

² Structure of the workfile

³ Main menu

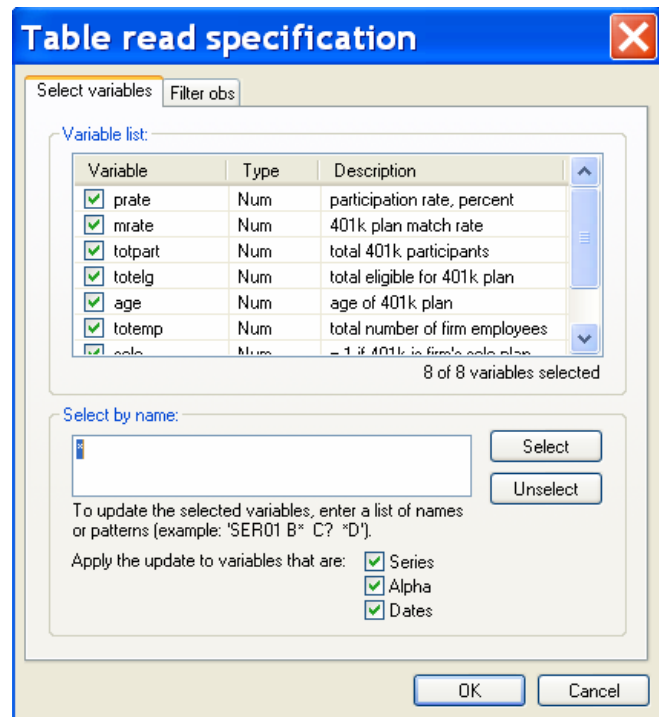
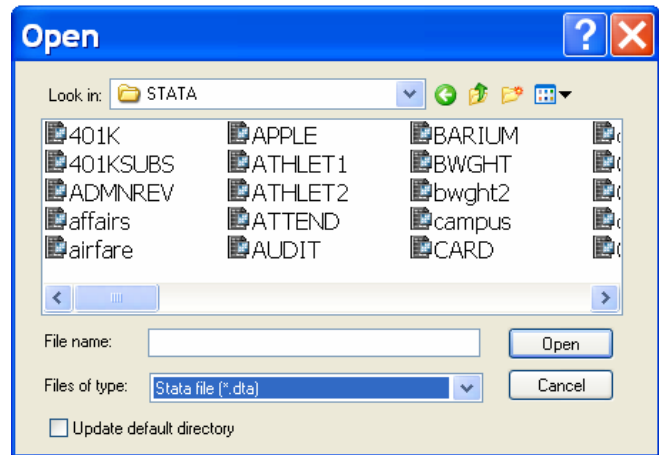
⁴ Nếu là dữ liệu năm, thì ở ô Frequency ta chọn Annual; ở các ô Start date và End date ta nhập năm bắt đầu và năm kết thúc của các chuỗi dữ liệu. Nếu dữ liệu là quý, thì ở ô Frequency ta chọn Quarterly; ở các ô Start date và End date ta nhập quý bắt đầu và quý kết thúc của các chuỗi dữ liệu. Ở đây ta có thể chọn một trong hai cách sau (ví dụ quý 2 năm 2005): 2005:2 hoặc 2005Q2. Nếu là dữ liệu tháng, thì ở ô Frequency ta chọn Monthly; ở các ô Start date và End date ta nhập tháng bắt đầu và tháng kết thúc của các chuỗi dữ liệu. Tương tự, ta có thể chọn một trong hai cách sau (ví dụ tháng 8 năm 2008): 2008:8 hoặc 2008M8. Các ô đặt tên là tùy chọn (đặt tên tập tin và tên trang), nhưng thông thường không cần thiết.

⁵ Sử dụng đối với loại dữ liệu chéo và ta chỉ cần nhập số quan sát của bộ dữ liệu (dataset) vào ô Observations là xong.

liệu thời gian; số quan sát đối với loại dữ liệu chéo; và tần suất, ngày bắt đầu, ngày kết thúc, và số quan sát tại mỗi thời điểm đối với loại dữ liệu bảng.

(ii) Tạo một tập tin bằng đọc từ một nguồn dữ liệu bên ngoài

Ta có thể mở trực tiếp một nguồn dữ liệu bên ngoài như cách mở một tập tin Eviews. Để mở một file bên ngoài, trước hết ta chọn **File/Open/Foreign Data as Workfile**¹, ... để đến hộp thoại **Open**, chọn **Files of type**, mở file cần chuyển sang tập tin Eviews, và thực hiện một số điều chỉnh nếu cần thiết. Xem ví dụ minh họa sau đây. Để mở và chuyển một tập tin nào đó sang Eviews, trước hết phải xác định thư mục thích hợp, rồi chọn tập tin (File name và Files of type) cần chuyển sang tập tin Eviews. Tuy nhiên, tập tin nguồn với định dạng khác nhau sẽ có một số khai báo riêng.



Đối với tập tin Stata. Khi chọn và mở tập tin (ví dụ Chapter2.1.dta trong thư mục data như trong hộp thoại), ta thấy xuất hiện hộp thoại **Table Read Specification**. Trong đó, ta chọn **Select** hoặc **Unselect** để chọn các biến cần thiết chuyển sang dạng dữ liệu Eviews thôi. Tuy nhiên, thông thường ta chọn tất cả các biến có sẵn theo mặc định của Eviews. Ngoài ra, ta cũng có thể định nghĩa lại bộ dữ liệu của mình thông qua chọn các điều kiện cần cho phù hợp mục tiêu nghiên cứu (ví dụ chỉ chọn các quan sát có age>10) bằng cách chọn **Filter Obs** và nhập điều kiện vào.

Đối với tập tin Text. Khi chọn và mở tập tin (ví dụ Chapter2.2.txt), ta thấy xuất hiện hộp thoại **ASCII Read**. Trong **Column specification** có ba lựa chọn: Delimiter ..., Fixed ..., và An explicit ... cho phép ta lựa chọn chiều rộng của các cột dữ liệu hiện trong tập tin. Tuy nhiên, thông thường Eviews sẽ mặc định ở dạng Delimiter ... Ở **Start date/header** ta thấy ô Skip lines cho phép ta lựa chọn bỏ các dòng đầu tiên (thường chỉ để lại dòng tên các biến), ví dụ ở đây ta chọn “2”. Điều này chỉ có ý nghĩa giúp ta dễ dàng kiểm tra dữ liệu chứ không cần thiết lắm. Mục **Row specification** cho phép ta xác định số quan sát trong một dòng (thông thường là 1). Mục này nói chung cũng không cần thiết. Sau đó ta chọn Next qua bước 2, và lại chọn Next để qua bước 3. Ở bước 3 ta có thể đặt lại tên biến bằng cách chọn biến đó và thay bằng tên biến

¹ Dĩ nhiên ta cũng có thể chọn **File/Open/Eviews Workfile** ... nhưng sẽ bất tiện một tí xíu là trong types of file sẽ mặc định loại tập tin Eviews thôi.

mong muốn (ví dụ biến “employment” đổi thành X₂). Ngoài ra, ta cũng có thể mô tả đặc điểm của biến đó (ví dụ đơn vị tính, ...). Cuối cùng ta chọn **Finish**.

ASCII Read - Step 1 of 3

Row specification
An observation in the file consists of a fixed number of:

Lines Lines per observation:

Characters Chars per observation:

Values Values per observation:
(ignores line breaks)

Column specification

Delimiter characters between values

Fixed width fields

An explicit format (to be provided)

Start of data/header
Skip lines:

Year	GDP	Employment	Fixed capital
1955	114043	8310	182113
1956	120410	8529	193749
1957	129187	8738	205192
1958	134705	8952	215130
1959	139960	9171	225021
1960	150511	9569	237026
1961	157897	9527	248897
1962	165286	9662	260661
1963	178491	10334	275466

ASCII Read - Step 3 of 3

Column headers

Header lines:

Header type:

Text representing NA

Column info
Click in preview to select column for editing

Name:

Description:

Data type:

Year	GDP	X2	Fixed capital
1955	114043	8310	182113
1956	120410	8529	193749
1957	129187	8738	205192
1958	134705	8952	215130
1959	139960	9171	225021
1960	150511	9569	237026
1961	157897	9527	248897
1962	165286	9662	260661

Thông thường nhất là sử dụng dữ liệu từ các tập tin Excel¹. Khi chọn và mở tập tin (ví dụ Chapter2.3.xls) Eviews sẽ thực hiện thông qua hai bước. Bước một, ta thấy xuất hiện hộp thoại **Spreadsheet Read** như sau:

Spreadsheet Read - Step 1 of 2

Cell Range

Predefined range

chapter2.3

Sheet: chapter2.3

Start cell: \$A\$1

End cell: \$E\$181

Custom range

chapter2.3!\$A\$1:\$E\$181

OBS	GDP	PR	M1	RS
1952:1	87.875	0.1975607	126.537	1.64
1952:2	88.125	0.1981673	127.506	1.677667
1952:3	89.625	0.2001787	129.385	1.828667
1952:4	92.875	0.2012459	128.512	1.923667
1953:1	94.625	0.2010517	130.587	2.047333
1953:2	95.55	0.2014442	130.341	2.202667
1953:3	95.425	0.2022359	131.389	2.021667
1953:4	94.175	0.2027231	129.891	1.486333
1954:1	94.075	0.2034164	130.173	1.083667

Cancel < Back Next > Finish

Bước hai, giống như bước 3 ở tập tin dạng Text, Eviews sẽ đưa ra các lựa chọn để đọc dữ liệu

và những thay đổi theo ý người sử dụng như đặt lại tên và nhãn của các biến. Tuy nhiên, trong hầu hết các trường hợp người sử dụng chỉ cần chọn

Finish để chấp nhận định dạng mặc định.

Spreadsheet Read - Step 2 of 2

Column headers

Header lines: 1

Header type: Names only

Clear Edited Column Info

Text representing NA

#N/A

Column info

Click in preview to select column for editing

Name: GDP

Description:

Data type: Number

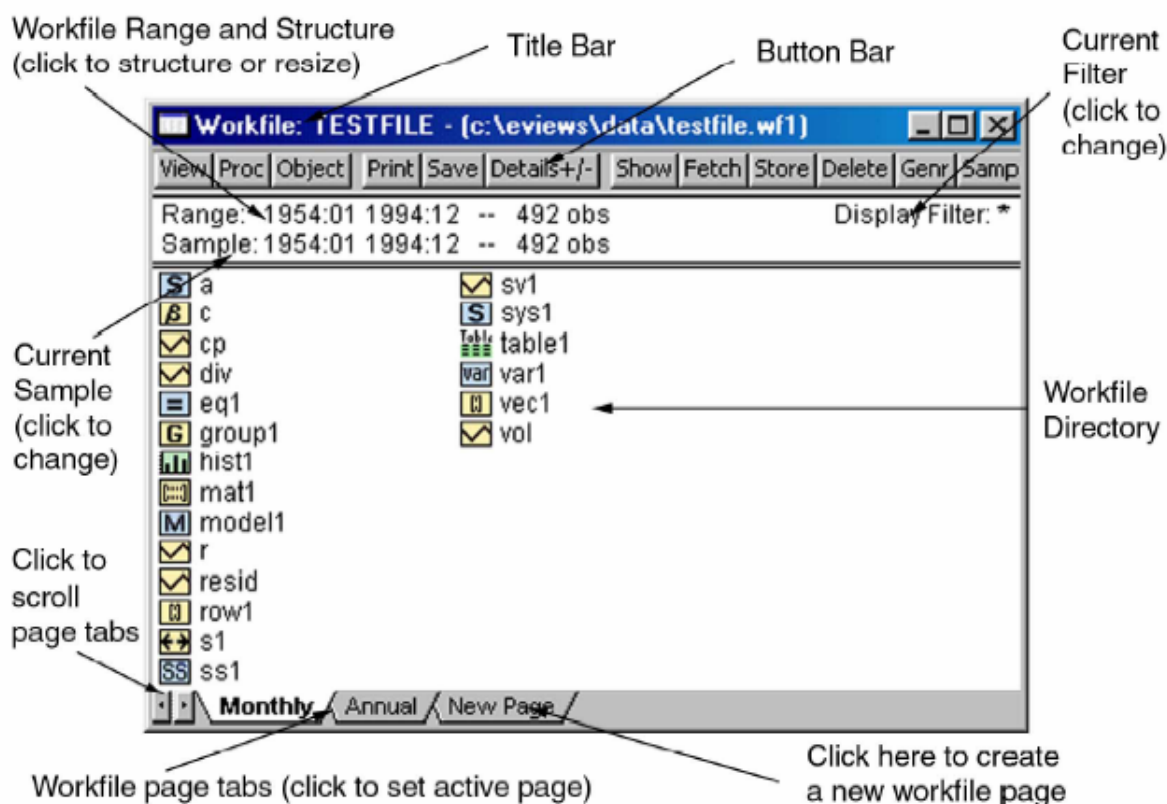
OBS	GDP	PR	M1	RS
1952:1	87.875	0.1975607	126.537	1.64
1952:2	88.125	0.1981673	127.506	1.677667
1952:3	89.625	0.2001787	129.385	1.828667
1952:4	92.875	0.2012459	128.512	1.923667
1953:1	94.625	0.2010517	130.587	2.047333
1953:2	95.55	0.2014442	130.341	2.202667
1953:3	95.425	0.2022359	131.389	2.021667
1953:4	94.175	0.2027231	129.891	1.486333

Cancel < Back Next > Finish

¹ Eviews 5 cho phép mở trực tiếp tất cả các tập tin dạng .xls, .raw, .txt, .dta, ... Ngày xưa ngày xưa khi chưa có Eviews 5, việc chuyển một tập tin từ Excel hay Text sang Eviews là một kỳ tích.

Nội dung cửa sổ tập tin của Eviews

Khi mở một tập tin làm việc của Eviews ta sẽ thấy một cửa sổ như sau:



Nguồn: Eviews 5 Users Guide, pp.52

Ta có thể trình bày dạng tóm tắt nội dung của tập tin Eviews bằng cách chọn **View/Statistics** và quay trở về thư mục gốc bằng cách chọn **View/Workfile Directory**.

Sau khi đã tạo một tập tin Eviews, ta nên lưu lại dưới định dạng Eviews bằng cách chọn **File/Save As ...** hay **File/Save ...** Eviews sẽ hiện ra hộp thoại **Saveas**, ta đặt tên cho tập tin đó, và chọn mức độ chính xác trong hộp thoại **Workfile Save**.

TRÌNH BÀY DỮ LIỆU

Khi đã có sẵn tập tin Eviews, ta có thể sử dụng các công cụ Eviews cơ bản để phân tích dữ liệu của từng chuỗi (sau đây cũng được gọi là biến¹) hay một nhóm các biến theo nhiều cách khác nhau.

Trình bày dữ liệu của một chuỗi²

Để xem nội dung của một biến nào đó, ví dụ M1³ trong tập tin Chapter2.3.wf1, ta nhấp đúp vào biểu tượng biến M1 trong cửa sổ của tập tin này, hay chọn **Quick/Show ...** trong thực đơn chính, nhập M1 và chọn OK. Eviews sẽ mở biến M1 và thể hiện dưới một dạng bảng tính mặc định.

¹ Variable

² Series statistics

³ ???

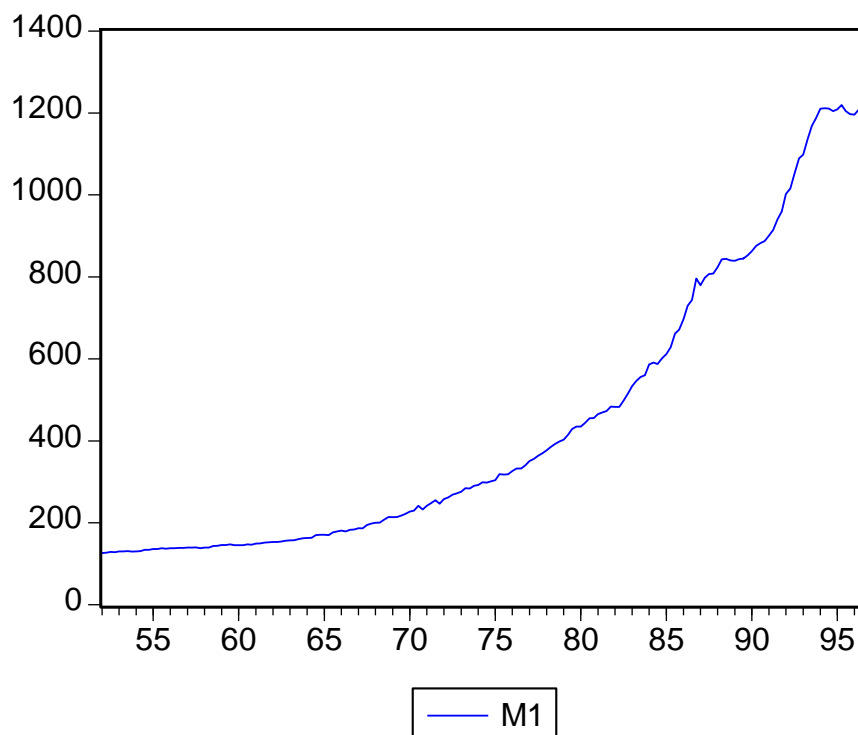
Series: M1 Workfile: CHAPTER2.3::Untitled	
M1	
Last updated: 07/23/08 - 12:37	
Imported from 'C:\Documents and Settings\User\My Documents\Kinh te luong\Sach Kinh te luong\...	
1952Q1	126.537
1952Q2	127.506
1952Q3	129.385
1952Q4	128.512
1953Q1	130.587
1953Q2	130.341
1953Q3	131.389
1953Q4	129.891
1954Q1	130.173
1954Q2	131.385
1954Q3	134.627
1954Q4	

- Đặt tên và nhãn của một chuỗi

Nhấp đúp vào thực đơn **Name**, Eviews sẽ hiển thị **Object Name** (tên đối tượng), trong đó có phần tên biến và nhãn của biến. Nếu biến có tên nhãn thì khi ta lập bảng hoặc vẽ đồ thị, thì trên bảng hay đồ thị sẽ hiển thị tên nhãn.

- Vẽ đồ thị

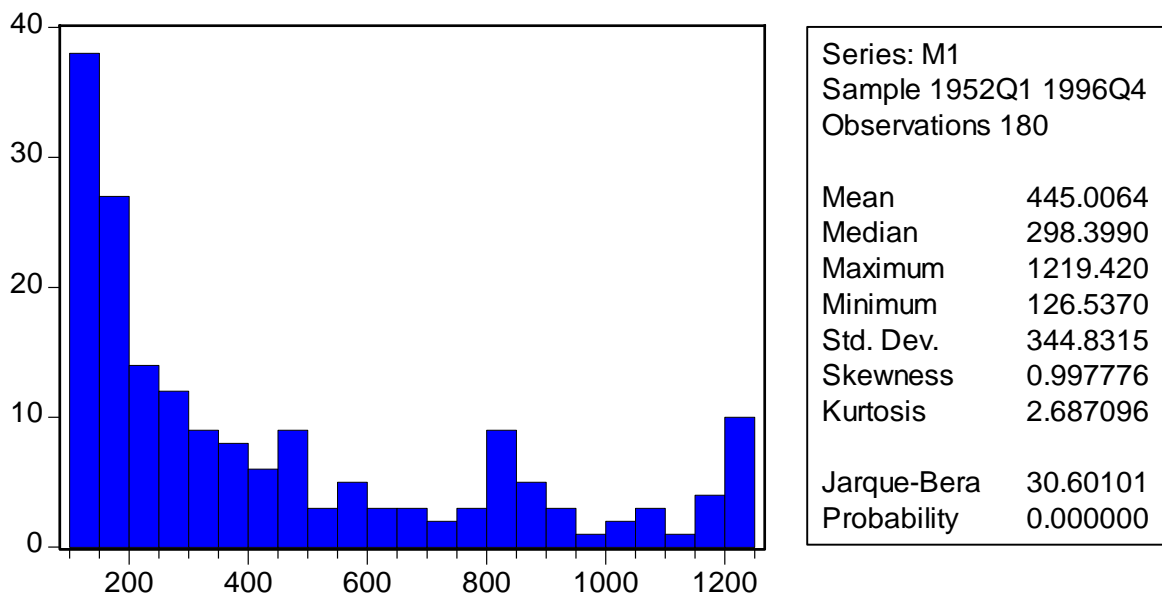
Có hai cách biểu diễn đồ thị dạng Line của biến M1. Thứ nhất, từ **Series M1** (chuỗi M1), ta chọn **View/Graph/Line**. Thứ hai, từ cửa sổ tập tin Chapter2.3.wf1 ta chọn **Quick/Graph/Line Graph**, ... rồi nhập tên biến M1, và OK. Lưu ý, để copy đồ thị ra word, ta chỉ cần click vào đồ thị và copy (có thể là Ctrl+C).



Để chỉnh sửa và biên tập đồ thị, ta chọn **Options** hay nhấp đúp vào đồ thị. Eviews cho phép thay đổi hình nền, khung đồ thị, dạng đường đồ thị, đổi trục¹, đặt tên nhãn đồ thị, thay đổi font chữ, ... Ngoài ra, Eviews cũng cho phép ta ghi chú dưới dạng text lên đồ thị, đặt tên đồ thị và lưu trong tập tin Eviews, hay có thể copy và dán dưới dạng văn bản. Ta cũng hay quan tâm đến các dạng biểu thị đồ thị bằng cách chọn **Template**, trong đó có nhiều sự lựa chọn rất thú vị. Nếu muốn lưu đồ thị (dạng một đối tượng trong tập tin Eviews), ta chọn **Object/Freeze Output**, rồi chọn Name để đặt tên đồ thị trong tập tin Eviews. Từ đồ thị, để trở lại bảng tính dữ liệu ta chọn **View/SpreadSheet**.

- Thống kê mô tả²

Để mô tả dữ liệu của một biến trên Eviews ta có thể làm như sau. Thứ nhất, từ **Series M1**, ta chọn **View/Descriptive Statistics/Histogram & Stats**. Thứ hai, từ cửa sổ tập tin Chapter2.3.wf1, ta chọn **Quick/Series Statistics/Histogram & Stats**, ... rồi nhập tên biến M1, và chọn OK. Lưu ý, để copy cả đồ thị và kết quả tóm tắt thống kê³, ta click vào đồ thị hoặc tóm tắt thống kê, và copy (có thể là Ctrl+C). Để lưu kết quả (đối tượng) này trong tập tin Eviews ta chọn **Freeze** (hoặc **Object/Freeze Output** và đặt tên. Ngoài ra, ta cũng có thể chỉnh sửa đồ thị như hướng dẫn ở trên.

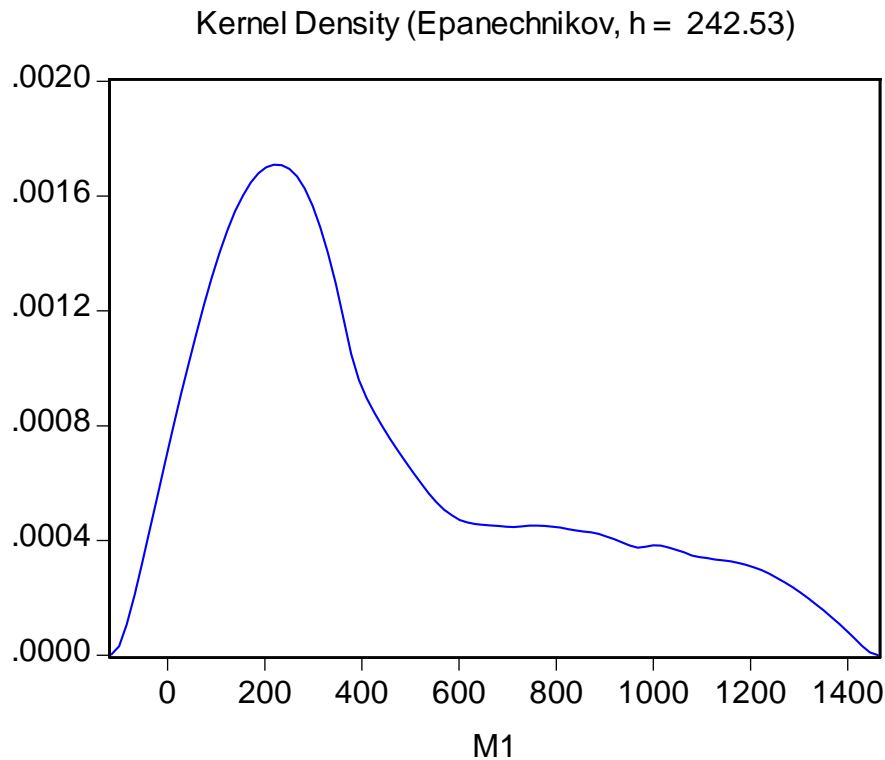


Ngoài ra, ta cũng có thể biểu diễn dưới dạng đồ thị phân phối xác suất của gián đồ phân phối histogram bằng cách chọn **View/Distribution/Kernel Density Graphs** và chọn OK.

¹ Eviews cũng cho phép vẽ đồ thị hệ trục kép (thường đối với các biến có đơn vị tính khác nhau như lượng cổ phiếu giao dịch và chỉ số giá chứng khoán)

² Có các loại thống kê như: thống kê mô tả (descriptive statistics), thống kê suy luận (statistical inference), thống kê toán (mathematical statistics), thống kê kinh tế (economic statistics), ... Tim đọc Hoàng Trọng, 2007, Thống kê ứng dụng trong kinh tế xã hội, NXB Thống kê.

³ Các thống kê này sẽ được trình bày chi tiết ở chương 3, hoặc có thể tìm đọc Hoàng Trọng, 2007. Đặc biệt, thống kê JB rất cần thiết cho việc kiểm định phân phối chuẩn của một biến số.



Trình bày dữ liệu của một nhóm các biến¹

- Mở và đặt tên nhóm

Để làm việc với một nhóm các biến có liên quan, ta chọn đồng thời các biến đó và chọn **Open as Group**, sau đó đặt tên nhóm bằng cách chọn **Name** để đặt tên nhóm và tên nhãn² của nhóm.

- Vẽ đồ thị

Cách vẽ đồ thị nhóm cũng tương tự vẽ đồ thị của từng chuỗi dữ liệu. Nếu vẽ đồ thị dạng Line giữa các chuỗi có thang đo khác nhau như GDP và M1, thì ta nên vẽ đồ thị hệ trục kép, trong đó các biến có thang đo tương tự nhau ta cho cùng hệ trục (trái hoặc phải). Các bước tiến hành như sau:

1. **Quick/Graph/Line, ... GDP M1, OK**
2. Chọn **Graph Options**, rồi chọn **Axes/Scales** để chuyển đồ thị của chuỗi M1 sang phải (**Right**)

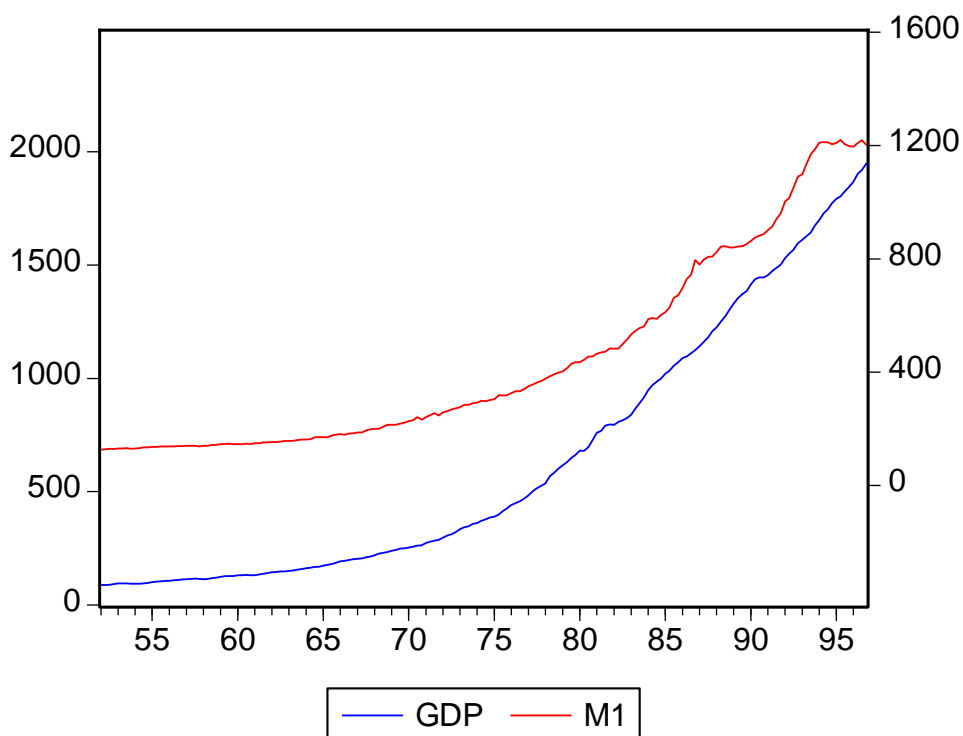
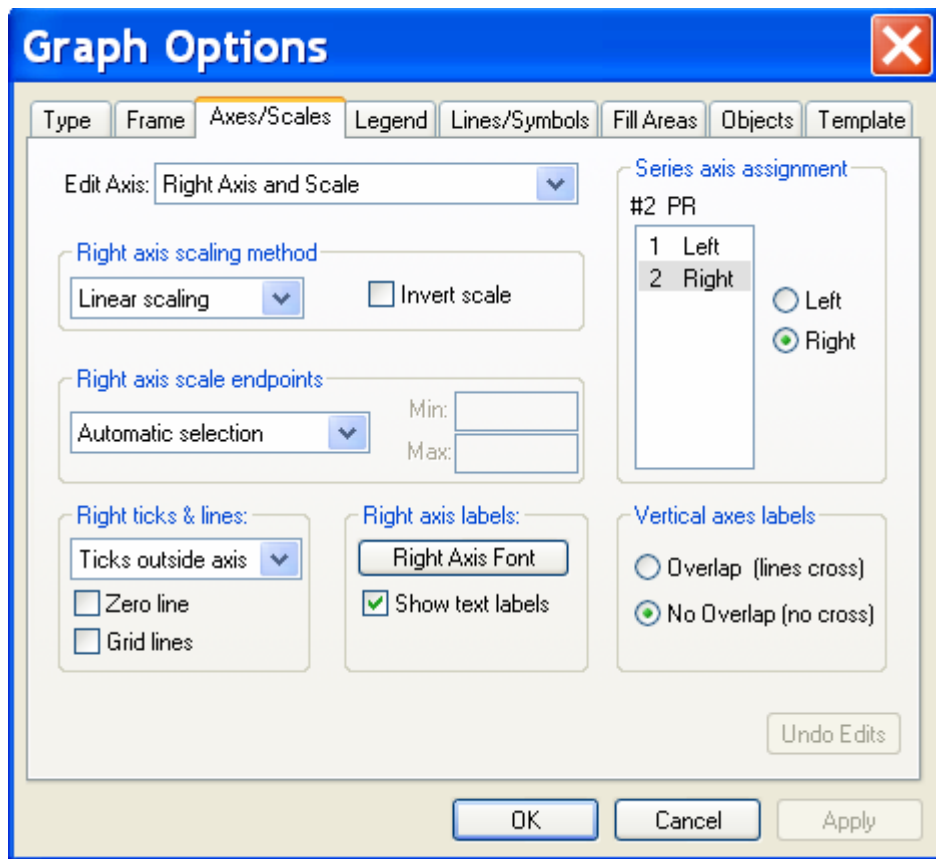
Ngoài ra, trong nhiều trường hợp người nghiên cứu có thể chọn dạng đồ thị thích hợp để biểu diễn mối quan hệ giữa các biến. Chẳng hạn, trường hợp hay gặp là trong mối quan hệ giữa lượng cổ phiếu giao dịch và chỉ số giá chứng khoán, thì người ta thường biểu diễn lượng cổ phiếu giao dịch bằng đồ thị dạng bar³ và chỉ số giá chứng khoán bằng đồ thị dạng line. Trong **Options** ta chọn **Type** và chọn dạng hỗn hợp (**Mixed**)⁴.

¹ Group statistics

² Label

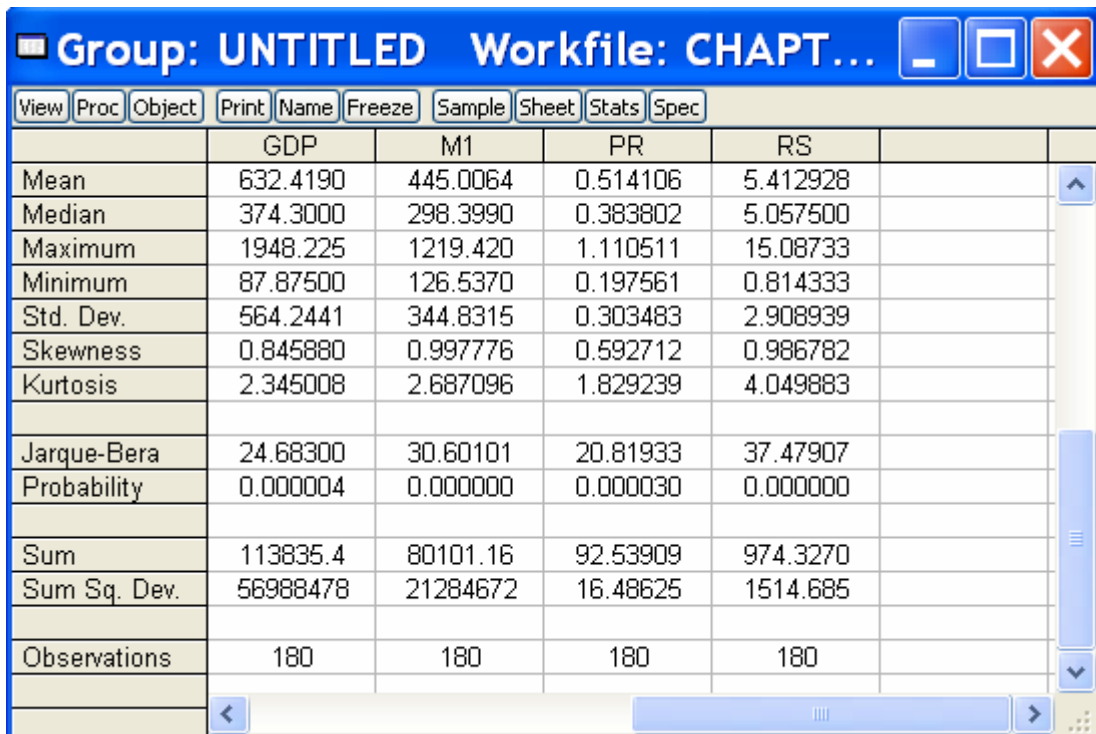
³ Thông thường là dạng cột (column), nhưng trường hợp này chưa gặp trong Eviews 5.

⁴ Điều quan trọng là 'đọc' được mối quan hệ giữa các biến trên đồ thị.



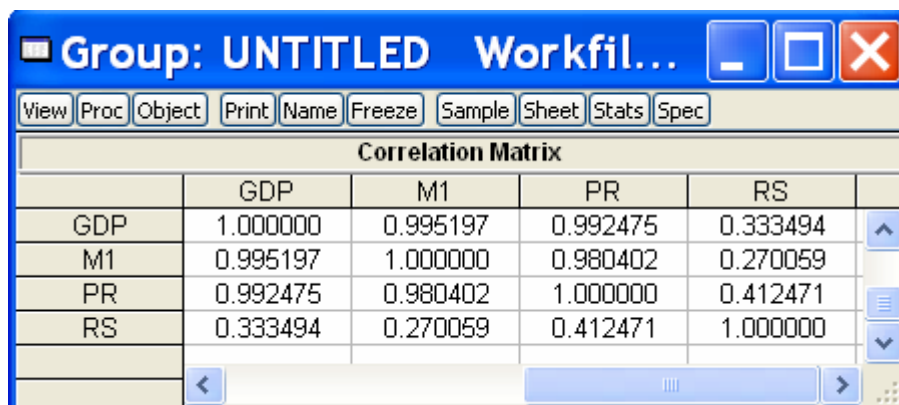
- Thống kê mô tả

Ta có thể đồng thời tạo ra một bảng thống kê mô tả nhiều biến khác nhau bằng cách chọn **View/Descriptive Stats/Individual Samples** hay **Quick/Group Statistics/Descriptive Statistics/Individual Samples**.



	GDP	M1	PR	RS
Mean	632.4190	445.0064	0.514106	5.412928
Median	374.3000	298.3990	0.383802	5.057500
Maximum	1948.225	1219.420	1.110511	15.08733
Minimum	87.87500	126.5370	0.197561	0.814333
Std. Dev.	564.2441	344.8315	0.303483	2.908939
Skewness	0.845880	0.997776	0.592712	0.986782
Kurtosis	2.345008	2.687096	1.829239	4.049883
Jarque-Bera	24.68300	30.60101	20.81933	37.47907
Probability	0.000004	0.000000	0.000030	0.000000
Sum	113835.4	80101.16	92.53909	974.3270
Sum Sq. Dev.	56988478	21284672	16.48625	1514.685
Observations	180	180	180	180

Ngoài ra, để xem xét mối quan hệ giữa từng cặp biến ta có thể tạo ma trận hệ số tương quan bằng cách chọn **View/Correlations/Pairwise Samples** hay **Quick/Group Statistics/Correlations**¹.



	GDP	M1	PR	RS
GDP	1.000000	0.995197	0.992475	0.333494
M1	0.995197	1.000000	0.980402	0.270059
PR	0.992475	0.980402	1.000000	0.412471
RS	0.333494	0.270059	0.412471	1.000000

CÁC VẤN ĐỀ CƠ BẢN VỀ ĐỐI TƯỢNG TRONG EIEWS

Đối tượng là tập hợp các thông tin và các thao tác có liên quan với nhau được nhóm lại thành một đơn vị nhằm mục đích tạo ra sự tiện lợi trong việc sử dụng. Hầu như tất cả các công việc thực hiện trên Eviews có liên quan đến sử dụng nhiều đối tượng khác nhau. Eviews giữ tất cả các đối tượng của nó trong các đối tượng chứa². Ta có thể hình dung đối tượng chứa giống như các tủ hồ sơ³ trong đó mỗi ngăn tủ là một đối tượng riêng. Đối tượng chứa quan trọng nhất trong Eviews là tập tin Eviews.

¹ Thống kê hệ số tương quan sẽ được trình bày ở chương 3

² Object container

³ Filing cabinet

Đối tượng là gì?

Thông tin trong Eviews được lưu trữ trong các đối tượng. Mỗi đối tượng gồm tập hợp các thông tin có liên quan nhau về một lĩnh vực phân tích nhất định. Ví dụ, đối tượng chuỗi là tập hợp các thông tin liên quan đến các quan sát của một biến số nhất định; đối tượng phương trình là tập hợp các thông tin liên quan đến mối quan hệ giữa một tập hợp các biến số. Lưu ý, một đối tượng không nhất thiết chỉ chứa đựng một thông tin duy nhất. Ví dụ, đối tượng phương trình ước lượng không chỉ chứa các hệ số ước lượng của phương trình¹, mà còn mô tả dạng mô hình, ma trận phương sai-hiệp phương sai² của các hệ số ước lượng, và nhiều thống kê khác nữa.

Dữ liệu của đối tượng

Mỗi đối tượng chứa đựng nhiều loại thông tin khác nhau. Ví dụ, các đối tượng chuỗi, ma trận, vectơ và tích vô hướng³ hầu như chỉ chứa thông tin số⁴. Ngược lại, các đối tượng phương trình và hệ thống chứa đựng các thông tin về dạng mô hình, và các kết quả ước lượng cũng như các tham chiếu về nguồn dữ liệu được sử dụng để ước lượng. Các đối tượng đồ thị và biểu bảng chứa cả các thông tin số, chữ, và định dạng. Do các đối tượng chứa đựng các loại dữ liệu khác nhau nên ta sẽ làm việc với các đối tượng khác nhau theo các cách khác nhau.

Các hiển thị đối tượng⁵

Có nhiều cách khác nhau để phân tích dữ liệu trong một đối tượng. Các hiển thị là các cửa sổ dạng biểu bảng hay đồ thị cung cấp cho ta nhiều cách khác nhau để xem xét dữ liệu trong một đối tượng. Ví dụ, một đối tượng chuỗi có thể có các cách hiển thị bảng tính, đồ thị đường thẳng, đồ thị thanh, thống kê và biểu đồ tần suất, giản đồ tự tương quan⁶, đồ thị phân phối, ... Một đối tượng phương trình có thể có các hiển thị dạng mô hình của phương trình, kết quả ước lượng, hiển thị giá trị thực-giá trị ước lượng-phần dư⁷ (kể cả các đồ thị). Một hiển thị hiệp phương sai chứa ma trận hiệp phương sai của các hệ số ước lượng, ...

Các hiển thị của một đối tượng được trình bày trong cửa sổ đối tượng⁸. Chỉ một cửa sổ có thể được mở cho mỗi đối tượng và tại một thời điểm mỗi cửa sổ chỉ trình bày một hiển thị duy nhất của một đối tượng. Dĩ nhiên, ta có thể thay đổi hiển thị của một đối tượng. Lưu ý, thay đổi hiển thị chỉ thay đổi định dạng của dữ liệu⁹ chứ không thể thay đổi dữ liệu trong đối tượng.

Các thủ tục của đối tượng¹⁰

Hầu hết các đối tượng của Eviews đều có các thủ tục. Giống như hiển thị, thủ tục thường trình bày các bảng biểu và đồ thị trong cửa sổ đối tượng. Tuy nhiên, khác hiển thị ở chỗ thủ tục có thể thay đổi dữ liệu trong bản thân đối tượng hoặc một đối tượng

¹ Estimated equation object

² Variance-Covariance matrix of the coefficient estimates

³ Scalar object

⁴ Numeric information

⁵ Object view

⁶ Correlogram

⁷ Actual-Fitted-Residual view

⁸ Object window






















⁹ Data display format

¹⁰ Object procedure/procs

khác. Nhiều thủ tục có thể tạo ra các đối tượng mới. Ví dụ, một đối tượng chuỗi có thể chứa các thủ tục làm trơn¹ hay điều chỉnh yếu tố mùa trong chuỗi thời gian và tạo ra một chuỗi mới chứa dữ liệu đã được làm trơn hay điều chỉnh. Đối tượng phương trình có các thủ tục tạo ra các chuỗi mới chứa phần dư, giá trị ước lượng, hay giá trị dự báo từ phương trình ước lượng.

Các loại đối tượng

Các đối tượng phổ biến nhất trong Eviews là chuỗi và phương trình. Tuy nhiên, có rất nhiều các đối tượng khác nhau và mỗi loại đối tượng có một chức năng nhất định. Hầu hết các đối tượng được biểu hiện bằng một biểu tượng² riêng. Dưới đây là các biểu tượng đối tượng cơ bản:

 Alpha	 Model	 Sym
 Coefficient Vector	 Pool	 System
 Equation	 Rowvector	 Table
 Graph	 Sample	 Text
 Group	 Scalar	 Valmap
 Logl	 Series	 VAR
 Matrix	 Sspace	 Vector

Các thao tác cơ bản về đối tượng³

• Tạo đối tượng

Để tạo một đối tượng trước hết ta phải mở tập tin chứa và của sổ tập tin chứa đang ở chế độ làm việc⁴, rồi chọn **Object/New Object** ở thực đơn chính. Khi đó ta thấy xuất hiện một hộp thoại **New Object**, sau đó chọn loại đối tượng.

• Chọn đối tượng

Cách dễ nhất để chọn đối tượng là chỉ vào-và-nhấp chuột. Hơn nữa, nút **View** trong thanh công cụ của tập tin Eviews có thể giúp ta chọn tất cả hoặc không chọn tất cả bằng cách chọn **Select All** hay **Deselect All**.

• Mở đối tượng

Sau khi đã chọn đối tượng hay một số đối tượng, chắc chắn ta sẽ muốn mở hoặc tạo ra một đối tượng mới chứa các đối tượng đã chọn. Thật đơn giản, ta chỉ cần nhấp đúp vào đối tượng đó. Nếu là một nhóm các đối tượng, ta phải chọn **View/Open as One Window ...**

¹ Smoothing

² Icon

³ Basic object operation

⁴ Active

- **Xem đối tượng¹**

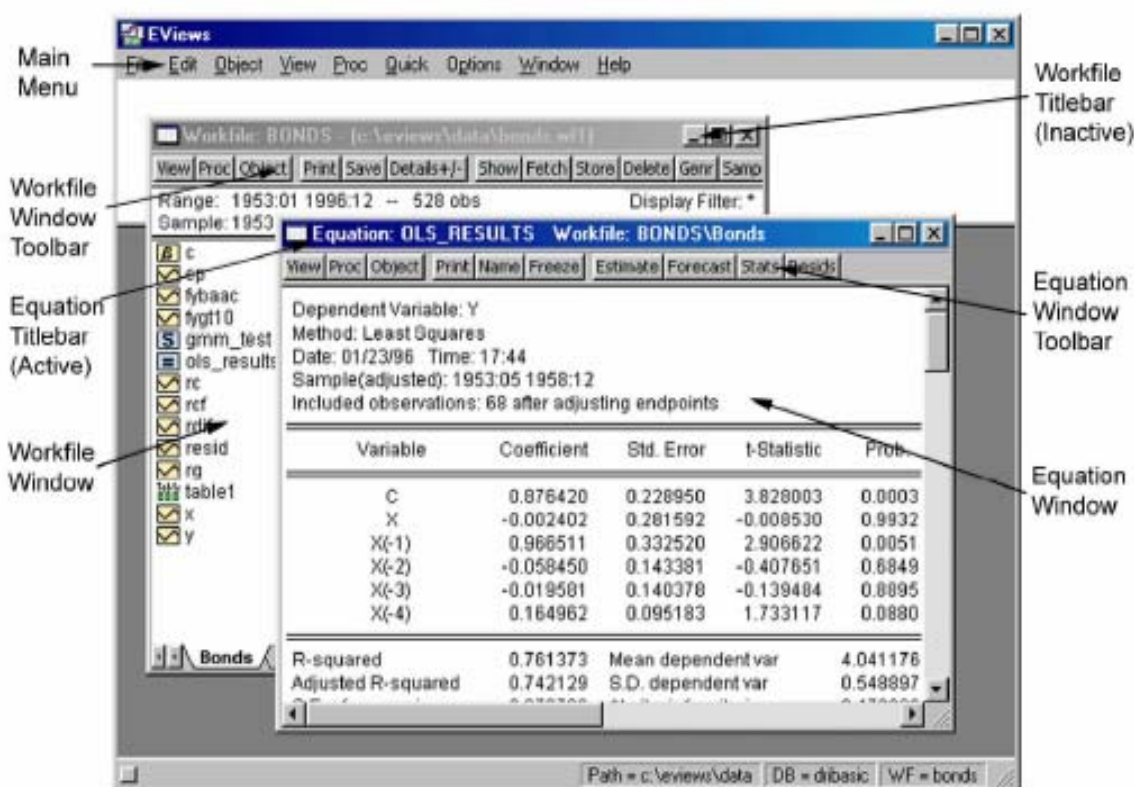
Một cách khác để chọn và mở đối tượng là chọn **Show** ở thanh công cụ² hay chọn **Quick/Show ...** từ thực đơn và nhập tên đối tượng vào hộp thoại. Nút Show cũng có thể được sử dụng để hiển thị các phương trình của các chuỗi.

- **Cửa sổ đối tượng**

Cửa sổ đối tượng là cửa sổ được hiển thị khi ta mở một đối tượng hay một chứa đối tượng. Một cửa sổ đối tượng sẽ chứa hoặc một hiển thị của đối tượng hoặc các kết quả của một thủ tục của đối tượng. Eviews cho phép mở cùng lúc nhiều cửa sổ đối tượng.

- **Các thành phần của một cửa sổ đối tượng**

Đây là minh họa cửa sổ phương trình từ kết quả hồi qui theo phương pháp OLS. Một số điểm cần lưu ý như sau:

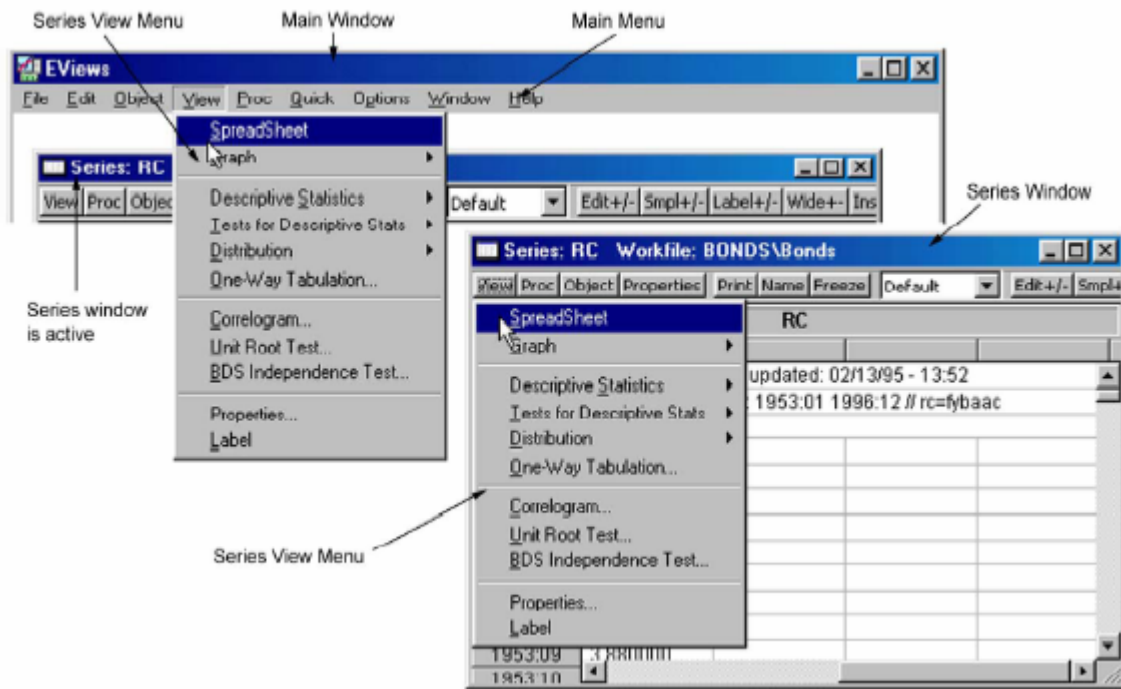


Thứ nhất, đây là một cửa sổ chuẩn vì ta có thể đóng, thay đổi kích cỡ, phóng to, thu nhỏ, và kéo lên xuống hay qua lại. Khi có nhiều cửa sổ khác đang mở, nếu ta muốn cửa sổ nào ở chế độ làm việc thì ta chỉ cần nhấp vào thanh tiêu đề hay bất kỳ đâu trong cửa sổ đó. Lưu ý, cửa sổ đang ở chế độ làm việc được biểu hiện với thanh tiêu đề có màu đậm. Thứ hai, thanh tiêu đề của cửa sổ đối tượng cho biết loại đối tượng, tên đối tượng, và tập tin chứa. Nếu đối tượng cũng chính là đối tượng chứa thì thông tin chứa được thay bằng thông tin thư mục. Thứ ba, trên đỉnh cửa sổ có một thanh công cụ chứa một số nút giúp ta dễ dàng làm việc.

¹ Show

² Toolbar

- **Các thực đơn và thanh công cụ của đối tượng**



Làm việc với đối tượng

- **Đặt tên và tên nhãn của đối tượng**

Các đối tượng có thể được đặt tên hoặc không được đặt tên. Khi ta đặt tên cho đối tượng, thì tên đối tượng sẽ xuất hiện trong thư mục của tập tin Eviews, và đối tượng sẽ được lưu như một phần của tập tin khi tập tin được lưu. Ta phải đặt tên đối tượng nếu muốn lưu lại các kết quả của đối tượng. Nếu ta không đặt tên, đối tượng sẽ được gọi là "UNTITLED". Các đối tượng không được đặt tên sẽ không được lưu cùng với tập tin, nên chúng sẽ bị xóa khi đóng tập tin.

Để đổi tên đối tượng, trước hết phải mở cửa sổ đối tượng, sau đó nhấp vào nút Name trên cửa sổ đối tượng và nhập tên (và tên nhãn) vào. Nếu có đặt tên nhãn thì tên nhãn sẽ xuất hiện trong các bảng biểu đồ thị, nếu không Eviews sẽ dùng tên đối tượng. Lưu ý, đây là nhóm đã mặc định và không được sử dụng cho tên đối tượng: ABS, ACOS, AND, AR, ASIN, C, CON, CNORM, COEF, COS, D, DLOG, DNORM, ELSE, ENDIF, EXP, LOG, LOGIT, LPT1, LPT2, MA, NA, NOT, NRND, OR, PDL, RESID, RND, SAR, SIN, SMA, SQR, và THEN.

- **Copy và dán đối tượng**

Có hai phương pháp tạo ra bản sao các thông tin chứa trong đối tượng: Copy và Freeze.

Nếu ta chọn **Object/Copy** từ thực đơn, Eviews sẽ tạo ra một đối tượng mới giống y như đối tượng gốc (dĩ nhiên phải khác tên). Ta cũng có thể copy đối tượng từ cửa sổ tập tin bằng cách chỉ ra đối tượng và chọn **Object/Copy Selected ...** sau đó xác định tên đích¹ cho đối tượng mới được copy.

Nếu ta chọn **Object/Freeze Output** hay chọn nút **Freeze** trên thanh công cụ của đối tượng, một đối tượng dạng bảng hay đồ thị được tạo ra giống y như hiển thị hiện hành của đối tượng gốc. **Freeze** hiển thị tạo ra một bản copy của hiển thị và tạo ra một đối tượng độc lập hoàn toàn. Tính chất cơ bản của việc **Freeze** một đối tượng là các bảng biểu và đồ thị được tạo ra có thể được chỉnh sửa cho mục đích trình bày hay báo cáo.

Ngoài ra, Eviews còn cho phép ta xóa, in ấn, lưu trữ, cập nhật, ... đối tượng.

QUẢN LÝ DỮ LIỆU

Có ba vấn đề cơ bản liên quan đến quản lý dữ² liệu trong Eviews là chuỗi, nhóm, và mẫu³. Mẫu là một số các quan sát trong tập tin Eviews được sử dụng cho các mục đích phân tích.

Một **chuỗi** trong Eviews là một số các quan sát về một biến bằng số, trong đó mỗi quan sát bao gồm ngày hoặc tên quan sát. Để tạo một chuỗi, từ cửa sổ tập tin Eviews ta chọn **Object/New Object/Series** và đặt tên cho chuỗi đó. Chuỗi mới được tạo ra chưa có giá trị bằng số và Eviews mặc định bằng ký hiệu NA. Ngoài ra, ta cũng có thể chọn **Quick/Generate Series** và trong hộp thoại **Enter Equation** ta đặt tên chuỗi mới, ví dụ Y=NA. Để định dạng một chuỗi dữ liệu trong Eviews, ví dụ M1, trước hết ta phải chọn và mở chuỗi M1, sau đó có thể chọn **Properties** để định dạng chuỗi, chọn **Name** để đặt tên và tên nhãn, chọn **Edit+/-** để điều chỉnh dữ liệu, chọn **InsDel** để bỏ hoặc chèn một quan sát ngay tại dấu nhắc hiện hành.

Khi làm việc đồng thời với nhiều chuỗi khác nhau như vẽ đồ thị, tính ma trận tương quan, kiểm định, ... ta thường tạo một nhóm riêng bằng cách chọn **Object/New Object ...** trong **Main Menu** hay **Workfile Toolbar**, chọn **Group**, nếu cần có thể đặt tên cho nhóm. Sau đó, ta phải nhập tên các chuỗi trong nhóm, ví dụ, GDP và M1, các chuỗi cách nhau bằng một khoảng trắng, cuối cùng chọn **OK**. Lưu ý, ta có thể đưa tên nhóm,



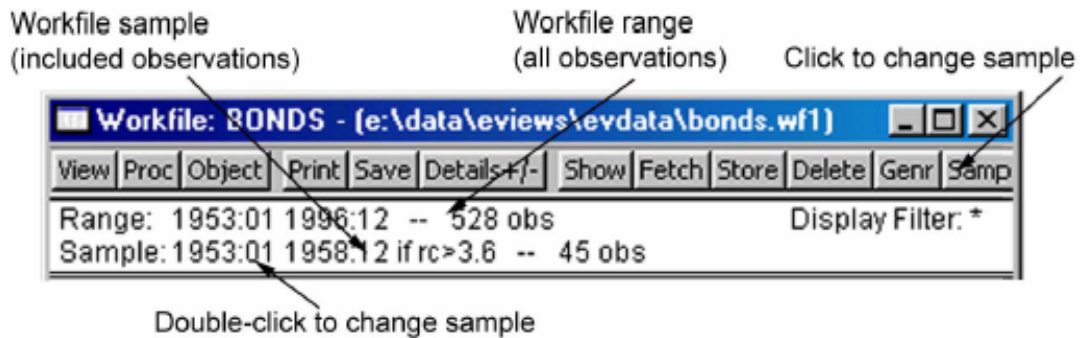
¹ Destination name, khác với original name

² Data handling

³ Được gọi chung là các đối tượng dữ liệu (data objects)

công thức, hay hàm số vào trong hộp thoại. Ngoài ra, ta có thể chọn **Quick/Show ...** rồi nhập tên các chuỗi, công thức, hàm số vào, rồi chọn Name để đặt tên nhóm.

Mẫu là một tập hợp các quan sát trong tập tin Eviews được sử dụng cho việc mô tả dữ liệu hoặc thực hiện các qui trình thống kê. Cần phân biệt hai khái niệm, **Workfile Range** - toàn bộ các quan sát sẵn có của một bộ dữ liệu, với **Workfile Sample** - chỉ các quan sát được sử dụng cho mục đích phân tích nhất định.



Có bốn cách xác định một mẫu tập tin Eviews: Thứ nhất, nhấp vào Sample trên Workfile Toolbar; Thứ hai, nhấp đúp vào mẫu hiện trong cửa sổ tập tin Eviews; Thứ ba, chọn **Proc/Sample ...** từ thực đơn chính của tập tin Eviews; Thứ tư, dùng lệnh `smpl` trong cửa sổ lệnh.

Để mở rộng một tập tin Eviews, ta chọn **Proc/Change Workfile Range ...**, nhập vào các ô **Start date** và **End date** khoảng thời gian mới. Khi đó, các quan sát mới của tất cả các chuỗi sẽ là NA.

Ngoài ra, Eviews cho phép ta có thể nhập và xuất dữ liệu từ các nguồn bên ngoài như Excel, Lotus, hay ASCII sang Eviews và ngược lại. Để biết chi tiết, ta có thể tham khảo chương 5, Eviews 5 Users Guide.

PHÉP TOÁN VÀ HÀM SỐ TRONG EIEWS

Phần này sẽ giới thiệu các nguyên tắc cơ bản khi sử dụng các phép tính toán học trong Eviews và hướng dẫn cách sử dụng các phép toán này khi làm việc với các dữ liệu chuỗi và nhóm.

Các phép toán số học¹

Các phép tính trong Eviews có thể là các toán tử² cho các phép toán số học thông thường. Trong Eviews, các toán tử cộng, trừ, nhân, chia, và lũy thừa lần lượt sẽ là +, -, *, /, và ^. Các phép so sánh trong Eviews được qui ước như sau: lớn hơn (>), nhỏ hơn (<), lớn hơn hoặc bằng (>=), nhỏ hơn hoặc bằng (<=), và bằng (=).

Các phép toán chuỗi

Eviews cho phép ta tính toán hoặc tạo một chuỗi mới từ một hoặc nhiều chuỗi đã có sẵn bằng các toán tử thông thường như trên. Ví dụ:

$$2*y+3$$

$$x/y+z$$

¹ Numeric expression

² Operator

Các hàm chuỗi

Hầu hết các hàm trong Eviews đều bắt đầu bằng ký hiệu @, ví dụ @mean(y) nghĩa là lấy giá trị trung bình của chuỗi y cho toàn bộ mẫu hiện hành. Có ba nhóm hàm chuỗi hay sử dụng trong Eviews: hàm toán (mathematical functions), hàm tập tin Eviews (workfile functions), và hàm dãy số (string functions). Để tìm hiểu thêm về các hàm này, ta có thể tham khảo ở **Help/Command & Programming Reference**, hoặc **Help/Quick Help Reference**, ở đây chỉ trình bày một số hàm hay sử dụng trong cuốn sách này.

Hàm giá trị tuyệt đối: @abs(x), abs(x)

Hàm mũ cơ số e hay antilog (e^x): @exp(x), exp(x)

Hàm nghịch đảo ($1/x$): @inv(x)

Hàm log tự nhiên ($\ln(x)$ hay $\log_e(x)$): @log(x), log(x)

Hàm căn bậc hai: @sqrt(x), sqr(x)

Hàm xu thế: @trend(base date), trong đó, base date chỉ thời điểm bắt đầu của chuỗi xu thế T (tại đó $T = 0$)

Biến trễ, tới, sai phân¹ và mùa vụ

Khi làm việc với dữ liệu chuỗi thời gian, ta thường xử lý dữ liệu bằng cách chuyển hóa sang dạng trễ, tới, sai phân, hoặc tạo thêm các biến giả mùa vụ.

- Biến trễ, tới và sai phân

Biến trễ một giai đoạn (x_{t-1}): x(-1)

Biến trễ k giai đoạn (x_{t-k}): x(-k)

Biến tới một giai đoạn (x_{t+1}): x(1)

Biến tới k giai đoạn (x_{t+k}): x(k)

Sai phân bậc một ($\Delta x = x_t - x_{t-1}$): d(x)

Sai phân bậc k ($\Delta^k x = x_t - x_{t-k}$): d(x,k)

Sai phân bậc một của biến trễ dạng log tự nhiên: dlog(x)

Trung bình trượt k giai đoạn: @movav(x,k)

Ngoài ra, ta có thể đồng thời kết hợp nhiều toán tử với nhau, ví dụ dlog(x), dlog(x,4), ...

- Biến giả mùa vụ

Tạo ra một biến giả theo quý có giá trị là 1 đối với quý 2 và giá trị là 0 đối với các quý khác: @seas(2)

Tạo ra một biến giả theo tháng có giá trị là 1 đối với tháng 2 và giá trị 0 đối với các tháng khác: @month(2)

¹ Lead: tới, Lag: trễ, và Difference: Sai phân

PHÂN TÍCH DỮ LIỆU VỚI EIEWS

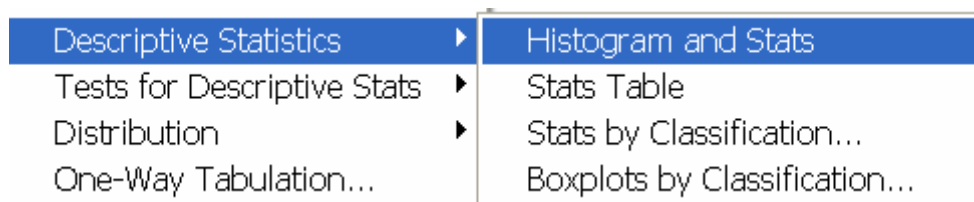
PHÂN TÍCH CHUỖI

Khi mở một chuỗi dữ liệu, nếu chọn View ta thấy xuất hiện một thực đơn dạng drop-down như hình bên cạnh. Thực đơn này chia thành bốn block riêng. Block thứ nhất mô tả dữ liệu dưới dạng bảng tính hay đồ thị. Block thứ hai và thứ ba cung cấp các công cụ thống kê cơ bản về chuỗi, trong đó block thứ ba chủ yếu sử dụng cho các chuỗi thời gian. Block thứ tư cho biết các thuộc tính, và tên nhãn của chuỗi.

Lưu ý rằng các phân phối xác suất của một biến ngẫu nhiên sẽ được trình bày một cách chi tiết ở chương 3.

Thống kê mô tả

Thống kê mô tả gồm có bốn nội dung: Đồ thị tần suất và thống kê (Histogram and Stats), Bảng thống kê (Stats Table), Thống kê theo nhóm (Stats by Classification ...), và Đồ thị hình hộp theo nhóm (Boxplots by Classification ...).



• Thống kê JB¹

Đây là một thống kê thường được sử dụng để kiểm định xem một biến có phân phối chuẩn hay không. Trong hồi qui tuyến tính cổ điển, thống kê này rất quan trọng cho việc kiểm định phần dư của mô hình hồi theo phương pháp OLS có phân phối chuẩn hay không.

Giả thiết H_0 : Chuỗi (biến) có phân phối chuẩn

$$\text{Công thức: } JB = \frac{N-k}{6} \left[S^2 + \frac{(K-3)^2}{4} \right] \quad (2.1)$$

Trong đó, S là độ nghiêng của phân phối, K là độ nhọn của phân phối, N là số quan sát, và k là số hệ số lượng được sử dụng để tạo ra chuỗi dữ liệu (N-k là bậc tự do)². Khi một chuỗi có phân phối chuẩn thì $S = 0$ và $K = 3$, nên $JB = 0$. Chúng ta sẽ biết ở chương 3 rằng thống kê JB có phân phối Chi bình phương với số bậc tự do là 2 ($\chi^2_{(2)}$)³. Xác suất được báo cáo kèm theo giá trị thống kê JB là xác suất mà thống kê

¹ Jarque-Bera

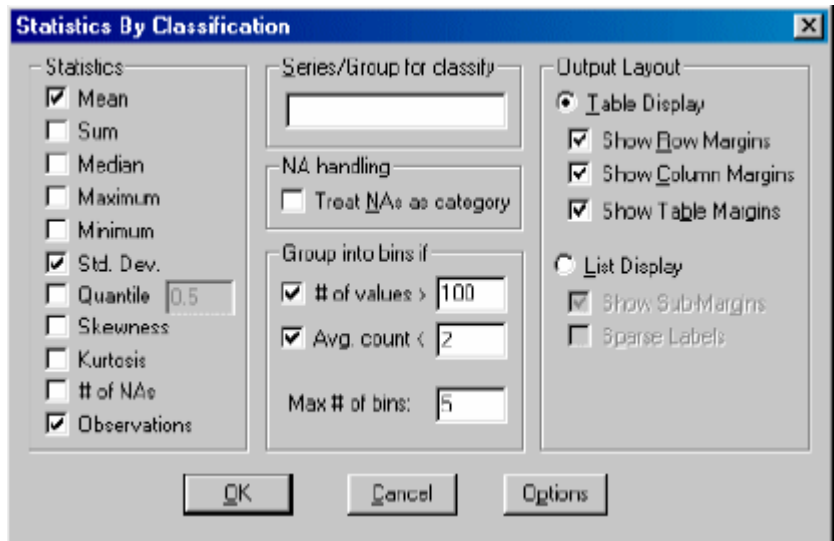
² Sẽ được giải thích ở chương 3 và 4. Có thể tham khảo thêm của Hoàng Trọng, 2007.

³ Sẽ được giải thích ở chương 4

JB lớn hơn giá trị quan sát (giá trị phê phán) dưới giả thiết không (H_0). Giá trị xác suất càng nhỏ thì khả năng bác bỏ giả thiết H_0 càng cao.

• **Thống kê theo nhóm**

Thống kê theo nhóm cho phép ta tính các thống kê mô tả của một chuỗi theo các phân nhóm khác nhau trong mẫu phân tích. Nếu ta chọn **View/Descriptive Statistics/Stats by Classification ...**, thì một hộp thoại sau đây sẽ xuất hiện:



Các lựa chọn ở **Statistics** bên trái cho phép ta chọn các tiêu chí thống kê muốn tính toán. Trong ô **Series/Group for classify** ta nhập tên chuỗi hay nhóm để xác định các phân nhóm. Nếu ta chọn nhiều chuỗi thì mỗi chuỗi cách nhau một khoảng trắng. Ở mục **Output Layout**, nếu ta chọn các **Margins** thì bảng kết quả có trình bày thống kê của tất cả các quan sát trong cùng một nhóm cũng như của toàn bộ mẫu phân tích. Ví dụ, sử dụng file **Chapter2.4.wf1** để thống kê mô tả biến LWAGE (log tự nhiên¹ của lương tuần) theo hai biến CONSTRUC (= 1 nếu làm việc trong ngành xây dựng và = 0 nếu làm trong các ngành khác) và MARRIED (= 1 nếu đã có gia đình và = 0 nếu chưa có gia đình). Kết quả thống kê biến LWAGE với bốn tiêu chí thống kê là trung bình, trung vị, lệch chuẩn, và số quan sát được trình bày như bảng bên cạnh. Nhìn vào bảng kết quả ta có thể so sánh có

Descriptive Statistics for LWAGE
 Categorized by values of CONSTRUC and MARRIED
 Date: 02/12/08 Time: 09:55
 Sample: 1 526
 Included observations: 526

	MARRIED	
	0	1
Mean	1.441258	1.736517
Median	1.316408	1.704748
Std. Dev.	0.509361	0.520818
Obs.	197	305
CONSTRUC		
0	1.506871	1.780785
1	1.504077	1.648659
	0.282824	0.496086
	9	15

sự khác biệt giữa các nhóm hay không. Tuy nhiên, để chắc chắn sự khác biệt đó có ý nghĩa về mặt thống kê hay không, ta cần dựa vào loại kiểm định thống kê thích hợp.

¹ Log tự nhiên được ký hiệu là ln, nhưng toán tử trong Eviews là log

Các kiểm định cho thống kê mô tả¹

• Kiểm định giả thiết đơn giản

Khi chọn View/Tests for Descriptive Stats/Simple Hypothesis Tests thì sẽ xuất hiện một hộp thoại như sau:

Trong hộp thoại này có ba loại kiểm định cơ bản là kiểm định trung bình, kiểm định phương sai, và kiểm định trung vị.

Kiểm định trung bình

Giả sử chuỗi X có giá trị trung bình mẫu là \bar{X} và giá trị trung bình tổng thể là μ . Giả sử ta có các giả thiết sau đây:

$$H_0: \mu = m$$

$$H_1: \mu \neq m$$

Nếu ta không biết giá trị độ lệch chuẩn của X, Eviews sẽ cho kết quả giá trị thống kê t² tính toán như sau:

$$t = \frac{\bar{X} - m}{s/\sqrt{N}} \quad (2.2)$$

Trong đó, s là độ lệch chuẩn của mẫu và N là số quan sát trong mẫu. Nếu X có phân phối chuẩn, thì với giả thiết H_0 , thống kê t sẽ theo phân phối t với bậc tự do là N-1.

Nếu ta biết giá trị độ lệch chuẩn của X, Eviews sẽ cho kết quả giá trị thống kê z³ tính toán như sau:

$$z = \frac{\bar{X} - m}{\sigma/\sqrt{N}} \quad (2.3)$$

Trong đó, σ là độ lệch chuẩn tổng thể của X. Nếu X có phân phối chuẩn với độ lệch chuẩn là σ , thì với giả thiết H_0 , thống kê z có phân phối chuẩn hóa.

Nếu giá trị t (z) tính toán lớn hơn giá trị t (z) phê phán ở một mức ý nghĩa xác định (α), ta bác bỏ giả thiết H_0 và ngược lại.

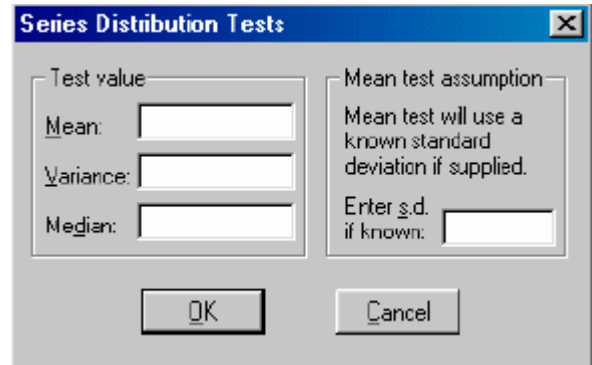
Kiểm định phương sai

Thực hiện kiểm định giả thiết H_0 cho rằng phương sai của chuỗi X bằng một giá trị σ^2 nhất định, ta có các giả thiết như sau:

$$H_0: \text{var}(X) = \sigma^2$$

$$H_1: \text{var}(X) \neq \sigma^2$$

Eviews sẽ cho kết quả thống kê chi bình phương tính toán như sau:



¹ Có thể tham khảo thêm trong Hoàng Trọng, 2007.

² Sẽ được giải thích ở chương 5

³ Khác biệt giữa thống kê z và thống kê t sẽ được giải thích ở chương 5

$$\chi^2 = \frac{(N-1)s^2}{\sigma^2} \quad (2.4)$$

Với giả thiết H_0 và giả định rằng X có phân phối chuẩn, thì thống kê chi bình phương sẽ theo phân phối chi bình phương với $N-1$ bậc tự do.

• **Kiểm định ngang bằng theo nhóm**

Đây là các kiểm định xem các giá trị trung bình, phương sai và trung vị ở các phân nhóm trong cùng một chuỗi có bằng nhau hay không. Khi chọn **View/Tests for Descriptive Stats/Equality Tests by Classification ...** sẽ thấy xuất hiện một hộp thoại như hình bên. Trước tiên ta phải chọn loại kiểm định: trung bình, phương sai, hay trung vị, sau đó chọn các phân nhóm muốn so sánh.

Xác định giả thiết:

Đối với kiểm định trung bình

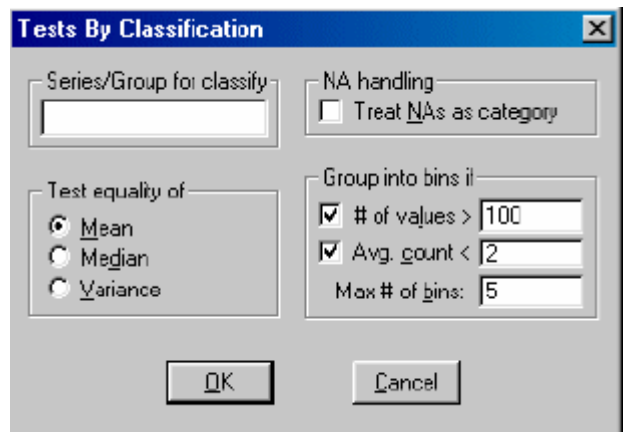
H_0 : Trung bình của các nhóm bằng nhau

H_1 : Trung bình của các nhóm khác nhau

Đối với kiểm định phương sai

H_0 : Phương sai của các nhóm bằng nhau

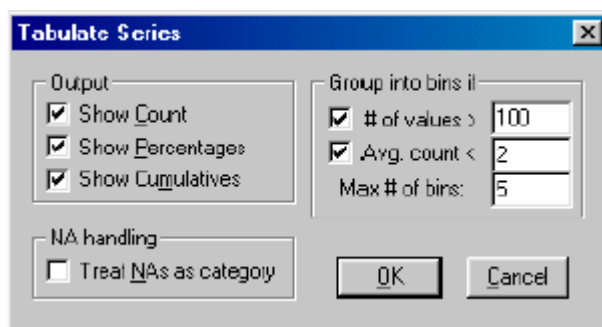
H_1 : Phương sai của các nhóm khác nhau



Để quyết định, ta so sánh giá trị thống kê F^1 tính toán với giá trị thống kê F quan sát (phê phán). Nếu giá trị thống kê F tính toán lớn hơn giá trị thống kê F quan sát tại một mức ý nghĩa nhất định, ta bác bỏ giả thiết H_0 và ngược lại.

Lập bảng tần suất một chiều

Để lập bảng tần suất một chiều ta chọn **View/One-Way Tabulation ...** và sẽ xuất hiện một hộp thoại như sau.



Biểu đồ tự tương quan

Mục đích của biểu đồ tự tương quan là giúp ta kiểm định xem một chuỗi thời gian dừng hay không dừng². Trong các mô hình dự báo chuỗi thời gian và dự báo bằng phương pháp hồi qui các chuỗi thời gian, thì việc các chuỗi thời gian dừng hay không

¹ Sẽ được giải thích ở chương 4 và 5

² Đây là một nội dung rất quan trọng khi phân tích chuỗi thời gian và đặc biệt có ý nghĩa rất lớn trong việc lựa chọn mô hình dự báo thích hợp trong các phương pháp dự báo định lượng với dữ liệu chuỗi thời gian. Nội dung này sẽ được trình bày chi tiết ở chương 14.

dùng có ý nghĩa rất quan trọng trong việc chọn mô hình dự báo thích hợp. Hai phương pháp kiểm định tính dừng thường được sử dụng là biểu đồ tự tương quan (dựa vào thống kê t và thống kê Q) và kiểm định nghiệm đơn vị¹ (dựa vào thống kê tau của Dickey-Fuller).

Biểu đồ tự tương quan là một đồ thị biểu diễn mối quan hệ giữa hệ số tự tương quan bậc k với độ trễ k tương ứng. Hệ số tự tương quan bậc k (ký hiệu là r_k) được xác định theo công thức sau đây:

$$r_k = \frac{\sum_{t=k+1}^n (Y_t - \bar{Y})(Y_{t-k} - \bar{Y})}{\sum_{t=1}^n (Y_t - \bar{Y})^2} \quad (2.5)$$

Trong đó, \bar{Y} là giá trị trung bình mẫu của chuỗi Y_t , k là độ trễ, n là số quan sát của mẫu. Có hai phương pháp kiểm định xem hệ số tự tương quan có ý nghĩa thống kê hay không: Thống kê t , và Thống kê Q^2 .

• Thống kê t

Gọi ρ_k là hệ số tự tương quan tổng thể (r_k là ước lượng không chệch của ρ_k), ta có các giả thiết sau đây:

$$H_0: \rho_k = 0$$

$$H_1: \rho_k \neq 0$$

Nếu một chuỗi thời gian ngẫu nhiên thì các hệ số tự tương quan là một biến ngẫu nhiên có phân phối chuẩn với trung bình là 0 và phương sai là $1/N$. Như vậy, với sai số chuẩn của hệ số tự tương quan $se(r_k)$ là $\sqrt{1/N}$, ta có thể xây dựng khoảng tin cậy cho ρ_k hoặc tìm được giá trị thống kê t tính toán ở một mức ý nghĩa xác định. Nếu ρ_k nằm ngoài khoảng tin cậy đó hoặc giá trị t tính toán lớn hơn giá trị t quan sát ta bác bỏ giả thiết H_0 .

• Thống kê Q

Hai cột cuối trong biểu đồ tự tương quan là thống kê Q của Ljung-Box và giá trị xác suất tương ứng. Thống kê Q kiểm định giả thiết đồng thời là tất cả các hệ số ρ_k cho tới một độ trễ đồng thời bằng không. Giá trị thống kê Q tính toán theo công thức sau đây:

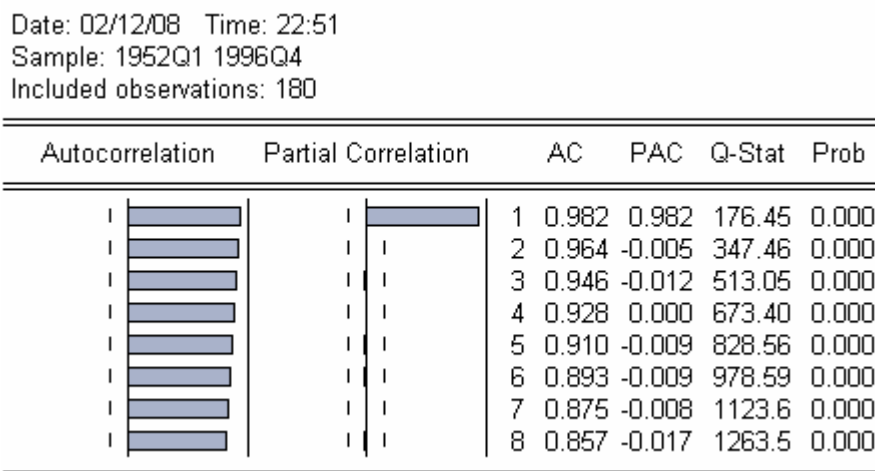
$$Q = n \sum_{k=1}^m \rho_k^2 \quad (2.6)$$

Với cỡ mẫu lớn, Q có phân phối theo Chi bình phương với bậc tự do bằng số độ trễ. Nếu giá trị thống kê Q tính toán lớn hơn giá trị thống kê Q quan sát ở một mức ý nghĩa xác định, ta bác bỏ giả thiết H_0 .

¹ Unit root test là một loại kiểm định rất mạnh và được sử dụng phổ biến. Nội dung này sẽ được trình bày chi tiết ở chương 14.

² Có thể tham khảo thêm trong John E.Hanke & Dean W.Wichern, (2005), Business Forecasting, 8th Edition.

Trong Eviews, ta lập biểu đồ tự tương quan bằng cách chọn **View/Correlogram ...**, xác định biểu đồ tự tương quan của chuỗi gốc hay chuỗi sai phân bậc một và bậc hai, và cuối cùng là xác định độ trễ k. Ví dụ, chuỗi GDP trong **Chapter2.3.xls** có biểu đồ tự tương quan như sau:



Dựa vào biểu đồ tự tương quan để xác định một chuỗi thời gian dừng hay không như sau. Có thể tóm tắt ý tưởng chính như sau. Nếu hệ số tự tương quan đầu tiên khác không nhưng các hệ số tự tương quan tiếp theo bằng không một cách có ý nghĩa thống kê, thì đó là một chuỗi dừng. Nếu một số hệ số tự tương quan khác không một cách có ý nghĩa thống kê thì đó là một chuỗi không dừng.

Kiểm định nghiệm đơn vị

Kiểm định nghiệm đơn vị là một kiểm định được sử dụng khá phổ biến để kiểm định một chuỗi thời gian dừng hay không dừng. Nội dung chi tiết phần kiểm định nghiệm đơn vị sẽ được trình bày ở chương 14 về các mô hình hồi qui chuỗi thời gian. Tuy nhiên, để hiểu qui trình kiểm định nghiệm đơn vị trên Eviews, ta nên xem qua một số ý tưởng cơ bản về mặt lý thuyết. Trước hết, cần lưu ý rằng có nhiều khái niệm chưa được học nên người đọc không nhất thiết phải hiểu ngay nội dung kiểm định nghiệm đơn vị ở chương này.

Giả sử ta có phương trình hồi qui tự tương quan như sau:

$$Y_t = \rho Y_{t-1} + u_t \quad (-1 \leq \rho \leq 1) \tag{2.7}$$

Ta có các giả thiết:

$$H_0: \rho = 1 \text{ (} Y_t \text{ là chuỗi không dừng)}$$

$$H_1: \rho < 1 \text{ (} Y_t \text{ là chuỗi dừng)}$$

Phương trình (2.7) tương đương với phương trình (2.8) sau đây:

$$\begin{aligned} Y_t - Y_{t-1} &= \rho Y_{t-1} - Y_{t-1} + u_t \\ &= (\rho - 1)Y_{t-1} + u_t \\ \Delta Y_t &= \delta Y_{t-1} + u_t \end{aligned} \tag{2.8}$$

Như vậy các giả thiết ở trên có thể được viết lại như sau:

$$H_0: \delta = 0 \text{ (} Y_t \text{ là chuỗi không dừng)}$$

$H_1: \delta < 0$ (Y_t là chuỗi dừng)

Dickey và Fuller cho rằng giá trị t ước lượng của hệ số Y_{t-1} sẽ theo phân phối xác suất τ (tau statistic, $\tau =$ giá trị δ ước lượng/sai số của hệ số δ). Kiểm định thống kê τ còn được gọi là kiểm định Dickey – Fuller (DF). Kiểm định DF được ước lượng với 3 hình thức:

- Khi Y_t là một bước ngẫu nhiên không có hằng số:

$$\Delta Y_t = \delta Y_{t-1} + u_t \quad (2.9)$$

- Khi Y_t là một bước ngẫu nhiên có hằng số:

$$\Delta Y_t = \beta_1 + \delta Y_{t-1} + u_t \quad (2.10)$$

- Khi Y_t là một bước ngẫu nhiên với hằng số xoay quanh một đường xu thế ngẫu nhiên:

$$\Delta Y_t = \beta_1 + \beta_2 \text{TIME} + \delta Y_{t-1} + u_t \quad (2.11)$$

Để kiểm định H_0 ta so sánh giá trị thống kê τ tính toán với giá trị thống kê τ tra bảng DF (các phần mềm kinh tế lượng đều cung cấp giá trị thống kê τ). Tuy nhiên, do có thể có hiện tượng tương quan chuỗi giữa các u_t do thiếu biến, nên người ta thường sử dụng kiểm định DF mở rộng là ADF (Augmented Dickey – Fuller Test). Kiểm định này được thực hiện bằng cách đưa thêm vào phương trình (2.11) các biến trễ của sai phân biến phụ thuộc ΔY_t :

$$\Delta Y_t = \beta_1 + \beta_2 \text{TIME} + \delta Y_{t-1} + \alpha_i \Sigma \Delta Y_{t-i} + \varepsilon_t \quad (2.12)$$

Để tiến hành kiểm định nghiệm đơn vị trên Eviews ta chọn **View/Unit Root Test ...**, sẽ xuất hiện hộp thoại **Unit Root Test**.

Ở lựa chọn **Test for unit root in**, chọn level nếu muốn kiểm định chuỗi gốc có phải là một chuỗi dừng hay không, chọn 1st difference nếu muốn kiểm định chuỗi sai phân bậc một có phải là một chuỗi dừng hay không. Ở lựa chọn **Include in test equation**, chọn intercept nếu dùng phương trình (2.10), chọn trend and intercept nếu dùng phương trình (2.11), chọn None nếu dùng phương trình (2.9), chọn trend and intercept và xác định độ trễ ở lựa chọn Lag length nếu dùng phương trình (2.12).

PHÂN TÍCH NHÓM

Khi mở một nhóm, nếu chọn View ta thấy xuất hiện thực đơn dạng drop-down như hình bên cạnh. Block thứ nhất cung cấp các cách khác nhau để mô tả dữ liệu trong nhóm. Block thứ hai trình bày các thống kê cơ bản. Block thứ ba chuyên về các thống kê của chuỗi thời gian. Block thứ tư là tên nhãn nhằm cung cấp các thông tin về nhóm. Trong phần này ta chỉ xem xét một số nội dung quan trọng thường được sử dụng trong kinh tế lượng.

Thống kê mô tả

Trong thống kê mô tả ta thấy có ba loại như sau: Common Sample, Individual Sample, và Boxplots. **Common Sample** chỉ tính các thống kê các quan sát có đầy đủ giá trị ở tất cả các chuỗi dữ liệu trong nhóm. **Individual Sample** tính các thống kê của các quan sát có đầy đủ giá trị ở mỗi chuỗi dữ liệu.

Kiểm định đồng liên kết¹

Chúng ta sẽ được biết ở chương 14 rằng khi hồi qui các chuỗi thời gian không dừng thường dẫn đến “kết quả hồi qui giả mạo”². Tuy nhiên, Engle và Granger³ (1987) cho rằng nếu kết hợp tuyến tính của các chuỗi thời gian không dừng có thể là một chuỗi dừng và các chuỗi thời gian không dừng đó được cho là đồng liên kết. Kết hợp tuyến tính dừng được gọi là phương trình đồng liên kết và có thể được giải thích như mối quan hệ cân bằng dài hạn giữa các biến. Nói cách khác, nếu phần dư trong mô hình hồi qui giữa các chuỗi thời gian không dừng là một chuỗi dừng, thì kết quả hồi qui là thực và thể hiện mối quan hệ cân bằng dài hạn giữa các biến trong mô hình. Mục đích của kiểm định đồng liên kết là xác định xem một nhóm các chuỗi không dừng có đồng liên kết hay không. Có hai cách kiểm định.

- **Kiểm định nghiệm đơn vị phần dư**

Giả sử GDP và M1 là hai chuỗi thời gian không dừng và ta có mô hình hồi qui như sau:

$$GDP_t = \beta_1 + \beta_2 M1_t + u_t \quad (2.13)$$

Nếu phần dư u_t là một chuỗi dừng thì kết quả hồi qui giữa GDP và M1 là “thực” và ta vẫn sử dụng một cách bình thường. Nói cách khác, GDP và M1 có quan hệ đồng liên kết và β_2 được gọi là hệ số hồi qui đồng liên kết. Các bước thực hiện trên Eviews như sau:

1) Ước lượng mô hình $GDP_t = \beta_1 + \beta_2 M1_t + u_t$

Group Members	
Spreadsheet	
Dated Data Table	
Graph ▶	
Multiple Graphs ▶	
<hr/>	
Descriptive Stats ▶	
Tests of Equality...	
N-Way Tabulation...	
Correlations ▶	
Covariances ▶	
Principal Components...	
<hr/>	
Correlogram (1) ...	
Cross Correlation (2) ...	
Cointegration Test...	
Unit Root Test...	
Granger Causality...	
<hr/>	
Label	

¹ Cointegration test

² Spurious regression

³ Đoạt giải Nobel kinh tế năm 2003

2) Kiểm định nghiệm đơn vị chuỗi u_t

• **Kiểm định đồng liên kết dựa trên phương pháp VAR của Johansen**

Eviews thực hiện kiểm định đồng liên kết trên cơ sở phương pháp luận VAR của Johansen (1991, 1995a). Lưu ý, kiểm định này chỉ có hiệu lực khi ta đang xét các chuỗi thời gian không dừng. Giả sử ta muốn kiểm định đồng liên kết giữa GDP và M1 trong **Chapter2.3.xls** theo phương pháp luận của Johansen, ta chọn **View/Cointegration Test ...** sẽ thấy xuất hiện một hộp thoại như sau:

Ở lựa chọn **Deterministic trend in data** có năm giả định về các chuỗi thời gian đang xem xét. Như sẽ được trình bày ở chương 14, một chuỗi thời gian có thể dừng sai phân hoặc dừng xu thế, trong đó có thể có xu thế xác định và xu thế ngẫu nhiên. Tương tự, các phương trình đồng liên kết có thể có hệ số cắt và xu thế xác định. Trên thực tế, trường hợp 1 và 5 ít khi được sử dụng. Nếu ta không chắc chắn về các giả định xi thế, ta nên chọn trường hợp 6. Nếu mô hình có các biến ngoại sinh thì ta đưa vào ô **exog variables**. Ngoài ra, ta có thể xác định độ trễ của biến phụ thuộc trong mô hình ở ô **Lag intervals** và mức ý nghĩa ở ô **MHM**. Kết quả kiểm định mối quan hệ đồng liên kết giữa GDP và M1 như bảng bên

Date: 02/13/08 Time: 13:14
 Sample (adjusted): 1953Q2 1996Q4
 Included observations: 175 after adjustments
 Trend assumption: Linear deterministic trend
 Series: GDP M1
 Lags interval (in first differences): 1 to 4

Unrestricted Cointegration Rank Test (Trace)

Hypothesized	Eigenvalue	Trace	0.05	
No. of CE(s)		Statistic	Critical Value	Prob.**
None *	0.174244	41.87126	15.49471	0.0000
At most 1 *	0.046684	8.366548	3.841466	0.0038

Trace test indicates 2 cointegrating eqn(s) at the 0.05 level
 * denotes rejection of the hypothesis at the 0.05 level
 **MacKinnon-Haug-Michelis (1999) p-values

ạnh. Ở đây có hai giả thiết H_0 : (i) “None”, nghĩa là không có đồng liên kết (đây là giả thiết ta quan tâm nhất); (ii) “At most 1”, nghĩa là có một mối quan hệ đồng liên kết. Lưu ý, tùy vào số biến trong mô hình (ví dụ k biến) mà ta có k-1 số phương trình đồng liên kết. Khi đó, ta có thêm số giả thiết về số phương trình đồng liên kết. Để quyết định bác bỏ hay chấp nhận giả thiết H_0 , ta so sánh giá trị “Trace Statistic” với giá trị phê phán (critical value) ở mức ý nghĩa xác định ở ô MHM (ở đây ta chọn là 5%).

- Nếu Trace Statistic < Critical Value, ta chấp nhận giả thiết H_0
- Nếu Trace Statistic > Critical Value, ta bác bỏ giả thiết H_0

Nhân quả Granger

Để kiểm định liệu có tồn tại mối quan hệ nhân quả Granger¹ giữa hai chuỗi thời gian Y và X trên Eviews, ta xây dựng hai phương trình sau:

$$Y_t = \alpha_0 + \alpha_1 Y_{t-1} + \dots + \alpha_k Y_{t-k} + \beta_1 X_{t-1} + \dots + \beta_l X_{t-l} + \varepsilon_t \quad (2.14)$$

$$X_t = \alpha_0 + \alpha_1 X_{t-1} + \dots + \alpha_l X_{t-l} + \beta_1 Y_{t-1} + \dots + \beta_k Y_{t-k} + \varepsilon_t \quad (2.15)$$

Để xem các biến trên của X có giải thích cho Y (X tác động nhân quả Granger lên Y) và các biến trễ của Y có giải thích cho X (Y tác động nhân quả Granger lên X) hay không ta kiểm định giả thiết sau đây cho mỗi phương trình:

$$H_0: \beta_1 = \beta_2 = \dots = \beta_l = 0 \quad (2.16)$$

Để kiểm định giả thiết đồng thời này, ta sử dụng thống kê F của kiểm định Wald² và cách quyết định như sau: Nếu giá trị thống kê F tính toán lớn hơn giá trị thống kê F phê phán ở một mức ý nghĩa xác định ta bác bỏ giả thiết H_0 và ngược lại. Có bốn khả năng như sau:

- Nhân quả Granger một chiều từ X sang Y nếu các biến trễ của X có tác động lên Y, nhưng các biến trễ của Y không có tác động lên X.
- Nhân quả Granger một chiều từ Y sang X nếu các biến trễ của Y có tác động lên X, nhưng các biến trễ của X không có tác động lên Y.
- Nhân quả Granger hai chiều giữa X và Y nếu các biến trễ của X có tác động lên Y và các biến trễ của Y có tác động lên X.
- Không có quan hệ nhân quả Granger giữa X và Y nếu các biến trễ của X không có tác động lên Y và các biến trễ của Y không có tác động lên X.

Để kiểm định nhân quả Granger trên Eviews ta chọn View/Granger Causality ... sẽ xuất hiện một hộp thoại về độ trễ tối ưu. Khi xác định độ trễ tối ưu và chọn OK, ta có kết quả như sau:

¹ Granger causality

² Kiểm định tập hợp ràng buộc tuyến tính sẽ được trình bày ở chương 7

Pairwise Granger Causality Tests
Date: 02/13/08 Time: 14:36
Sample: 1952Q1 1996Q4
Lags: 2

Null Hypothesis:	Obs	F-Statistic	Probability
M1 does not Granger Cause GDP	178	0.47794	0.62088
GDP does not Granger Cause M1		4.80380	0.00932

Lưu ý, các độ trễ của X và Y có thể khác nhau và có thể được xác định bằng một số tiêu chí thống kê khác nhau.

XÂY DỰNG HÀM KINH TẾ LƯỢNG TRÊN EIEWS

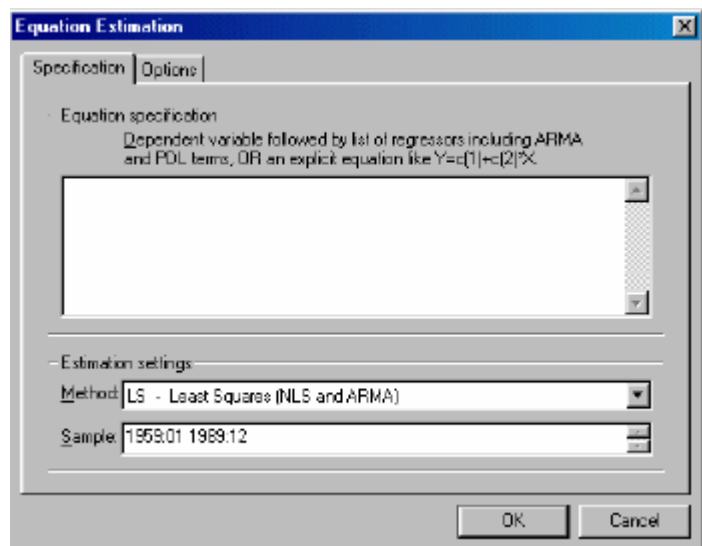
Trong tài liệu này ta chỉ xét phân tích hồi qui đơn phương trình. Phần này trình bày các kỹ thuật phân tích hồi qui cơ bản như xác định và ước lượng một mô hình hồi qui, kiểm định giả thiết, và sử dụng kết quả ước lượng cho các mục đích dự báo.

ĐỐI TƯỢNG PHƯƠNG TRÌNH

Ước lượng hồi qui đơn phương trình trên Eviews được thực hiện bằng cách sử dụng đối tượng phương trình. Để tạo ra một đối tượng phương trình ta chọn **Object/New Object ... /Equation** hay **Quick/Estimate Equation ...** từ thực đơn chính, hay đơn giản chỉ cần đánh **equation** trong cửa sổ lệnh. Kế tiếp, ta sẽ xác định dạng phương trình trong hộp soạn thảo **Specification** của hộp thoại **Equation Estimation** và chọn phương pháp ước lượng ở ô **Method**. Các kết quả ước lượng được lưu trữ như một phần của đối tượng phương trình.

Xác định phương trình hồi qui

Khi tạo ra một đối tượng phương trình sẽ thấy xuất hiện một hộp thoại **Equation Estimation** và ta cần xác định ba việc sau: dạng phương trình, phương pháp ước lượng, và mẫu được sử dụng để ước lượng. Trong hộp soạn thảo dạng phương trình ta nhập các biến phụ thuộc và giải thích theo thứ tự từ trái qua phải và lưu ý xác định dạng hàm. Có hai cách xác định dạng phương trình ước lượng: liệt kê các biến và công thức. Phương pháp liệt kê dễ hơn nhưng chỉ có thể sử dụng giới hạn ở các dạng mô hình tuyến tính. Phương pháp công thức tổng quát hơn và phải được sử dụng để xác định các dạng mô hình phi tuyến và các mô hình có ràng buộc tham số.



Xác định phương trình theo phương pháp liệt kê

Cách đơn giản nhất để xác định một phương trình tuyến tính là liệt kê các biến trong phương trình. Trước hết, nhập tên biến phụ thuộc hay công thức của biến phụ thuộc,

sau đó nhập tên các biến giải thích. Ví dụ, sử dụng **Chapter2.3.xls** để xác định phương trình hồi qui GDP theo cung tiền M1, ta nhập vào hộp thoại soạn thảo dạng phương trình như sau:

$$\text{GDP c M1} \quad (2.17)$$

Lưu ý có hiện diện của chuỗi C trong danh sách các biến giải thích. Đây là một chuỗi mặc định sẵn trong Eviews được dùng để xác định hằng số trong phương trình hồi qui. Eviews không tự động đưa hằng số vào phương trình hồi qui vì tùy thuộc vào mô hình có hệ số cắt hay không nên ta phải đưa vào khi xác định phương trình hồi qui. C là một đối tượng đã được xác định trước trong bất kỳ một tập tin Eview nào. Đây là một vector hệ số mặc định – khi ta xác định phương trình bằng cách liệt kê tên các biến, Eviews sẽ lưu giữ các hệ số ước lượng trong vector này theo thứ tự xuất hiện trong danh sách các biến. Trong ví dụ trên, hằng số sẽ được lưu trong C(1) và hệ số của M1 sẽ được lưu trong C(2).

Nếu mô hình có biến trễ một giai đoạn của biến phụ thuộc thì ta liệt kê các biến trong hộp thoại soạn thảo này như sau:

$$\text{GDP GDP(-1) c M1} \quad (2.18)$$

Như vậy hệ số của biến trễ biến GDP là C(1), hệ số của hằng số là C(2), và hệ số của M1 là C(3).

Nếu mô hình có nhiều biến trễ liên tục của biến GDP thì thay vì phải nhập từng biến trễ GDP(-1) GDP(-2) GDP(-3) GDP(-4), Eviews cho phép thực hiện như sau:

$$\text{GDP GDP(1 to 4) c M1} \quad (2.19)$$

Tuy nhiên, nếu ta không đưa số 1 và dấu ngoặc đơn thì Eviews sẽ hiểu đó là số 0. Ví dụ:

$$\text{GDP c M1(to 2) M1(-4)} \quad (2.20)$$

Thì Eviews sẽ hiểu ta hồi qui GDP theo hằng số C, M1, M1(-1), M1(-2), và M1(-4). Ngoài ra, ta cũng có thể đưa các chuỗi điều chỉnh vào nhóm các biến giải thích. Ví dụ ta hồi qui GDP theo hằng số, biến trễ của GDP, và biến trung bình di động hai giai đoạn của M1 như sau:

$$\text{GDP GDP(-1) c ((M1+M1(-1))/2)} \quad (2.21)$$

Xác định phương trình theo phương pháp công thức

Một công thức phương trình trong Eviews là một biểu thức toán về các biến và hệ số. Để xác định một phương trình bằng công thức, đơn giản là ta nhập biểu thức vào hộp thoại soạn thảo. Ví dụ, hồi qui mô hình dạng log tự nhiên như sau:

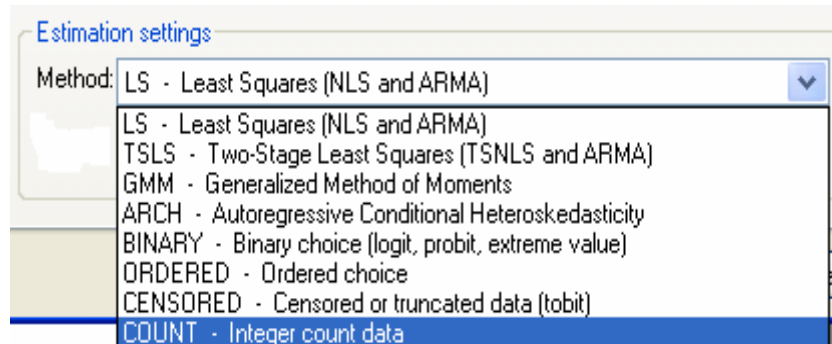
$$\log(\text{GDP}) \text{ c } \log(\text{GDP}(-1)) \log(\text{M1}) \quad (2.22)$$

Hai lý do chủ yếu ta phải sử dụng phương pháp công thức này là ước lượng các mô hình ràng buộc và phi tuyến.

Ước lượng một phương hồi qui

Phương pháp ước lượng

Sau khi đã xác định phương trình, ta cần chọn phương pháp ước lượng bằng cách nhấp vào **Method** và sẽ thấy xuất hiện một hộp thoại dạng drop-down liệt kê các phương pháp ước lượng.



Phương pháp sử dụng phổ biến nhất đối với hồi qui đơn phương trình là phương pháp bình phương bé nhất¹. Trong chương trình kinh tế lượng căn bản của cuốn sách này, ta chỉ sử dụng hai phương pháp là LS - Least Squares² và BINARY – Binary choice³. Hai phương pháp này sẽ được trình bày chi tiết vào các chương sau.

Mẫu ước lượng

Ta nên xác định mẫu sử dụng cho việc ước lượng mô hình. Theo mặc định, Eviews đưa ra mẫu của tập tin Eviews hiện hành, nhưng ta có thể thay đổi mẫu theo mục đích ước lượng bằng cách nhấp vào hộp thoại **Sample**. Thay đổi mẫu ở đây không ảnh hưởng gì đến mẫu của tập tin Eviews hiện hành. Nếu có quan sát không có giá trị⁴, Eviews tạm thời điều chỉnh mẫu ước lượng để loại bỏ các quan sát đó ra khỏi mẫu phân tích.

```

Dependent Variable: Y
Method: Least Squares
Date: 08/19/97 Time: 10:24
Sample(adjusted): 1959:01 1989:12
Included observations: 340
Excluded observations: 32 after adjusting endpoints

```

Ngoài ra, nếu trong mô hình có các biến trễ hay biến điều chỉnh thì Eviews cũng điều chỉnh số mẫu ước lượng.

¹ Least squares/Ordinary least squares

² Kể cả phương pháp WLS (Weighted least squares) và GLS (Generalized least squares)

³ Hai loại mô hình sẽ được trình bày ở chương 15 là Logit và Probit

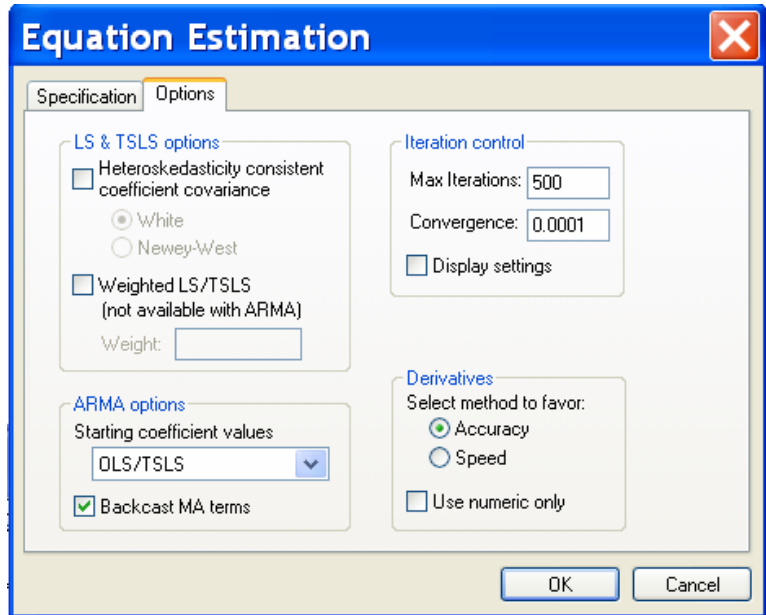
⁴ Missing value

Các lựa chọn ước lượng

Khi chọn **Options** ta sẽ thấy xuất hiện hộp thoại **Equation Estimation**. Các nội dung trong phần lựa chọn ước lượng như Heteroskedasticity consistent coefficient covariance và Weighted LS/TSLS sẽ được trình bày chi tiết ở chương 11 và 12.

Kết quả ước lượng

Sau khi đã hoàn thành các bước trên ta chọn OK trong hộp thoại **Equation Estimation**,



Eviews sẽ hiển thị cửa sổ phương trình về hiển thị kết quả ước lượng. Trong kết quả ước lượng của Eviews gồm ba phần chính: Tóm tắt các đặc điểm của

mô hình hồi qui (biến phụ thuộc, phương pháp ước lượng, thời điểm thực hiện ước lượng, mẫu ước lượng, và số quan sát được sử dụng cho ước lượng kết quả); Kết quả hệ số (tên các biến giải thích, giá trị ước lượng các hệ số hồi qui, sai số chuẩn, thống kê t, và giá trị xác suất); và Tóm tắt thống kê (hệ số xác định

Dependent Variable: GDP
 Method: Least Squares
 Date: 07/24/08 Time: 11:50
 Sample (adjusted): 1952Q2 1996Q4
 Included observations: 179 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	3.375294	1.081021	3.122321	0.0021
GDP(-1)	1.023484	0.008443	121.2196	0.0000
M1	-0.017148	0.013636	-1.257562	0.2102
R-squared	0.999886	Mean dependent var		635.4611
Adjusted R-squared	0.999885	S.D. dependent var		564.3446
S.E. of regression	6.056304	Akaike info criterion		6.456694
Sum squared resid	6455.471	Schwarz criterion		6.510114
Log likelihood	-574.8741	F-statistic		772705.5
Durbin-Watson stat	1.321357	Prob(F-statistic)		0.000000

R^2 , R^2 điều chỉnh, sai số chuẩn của hồi qui, tổng bình phương phần dư (RSS), thống kê d Durbin-Watson, AIC, SIC, thống kê F, ...). Sau khi học xong chương trình kinh tế lượng căn bản ít nhất ta sẽ hiểu một cách hệ thống tất cả các thông tin trong bảng kết quả ước lượng này.

Phân tích kết quả hồi qui

Các hiển thị của phương trình hồi qui

Các hình thức biểu diễn phương trình hồi qui¹.

Biểu diễn phương trình hồi qui dưới ba hình thức: hình thức lệnh ước lượng trong màn hình lệnh của Eviews, hình thức phương trình đại số dưới dạng các ký hiệu, và hình thức phương trình với các giá trị ước lượng các hệ số.

Estimation Command:

=====

LS GDP GDP(-1) C M1

Estimation Equation:

=====

GDP = C(1)*GDP(-1) + C(2) + C(3)*M1

Substituted Coefficients:

=====

GDP = 1.02348388*GDP(-1) + 3.375294154 - 0.0171476024*M1

Kết quả ước lượng. Như đã trình bày ở trên.

Giá trị thực tế, giá trị ước lượng, và phần dư. Có các cách trình bày như sau:

Actual,Fitted,Residual	▶	Actual,Fitted,Residual Table
ARMA Structure...		Actual,Fitted,Residual Graph
Gradients and Derivatives	▶	Residual Graph
Covariance Matrix		Standardized Residual Graph

Các kiểm định về hệ số hồi qui, phần dư, và sự ổn định. Đây là các nội dung rất quan trọng trong phân tích hồi qui và sẽ được trình bày ở các phần sau.

Các thủ tục của phương trình hồi qui

Có nhiều thủ tục liên quan đến kết quả ước lượng phương trình hồi qui, nhưng hai thủ tục hay sử dụng là **Specify/Estimate** ... nghĩa là giúp quay lại hộp thoại **Equation Estimation** một cách nhanh chóng nếu ta có bất kỳ điều chỉnh nào về mô hình hồi qui, và **Forecast** ... nghĩa là nếu mô hình ước lượng tốt ta có thể dùng cho mục đích dự báo và thủ tục này giúp ta xác định được giá trị dự báo điểm, xây dựng đồ thị khoảng dự báo, và các tiêu chính thống kê về độ chính xác của dự báo.

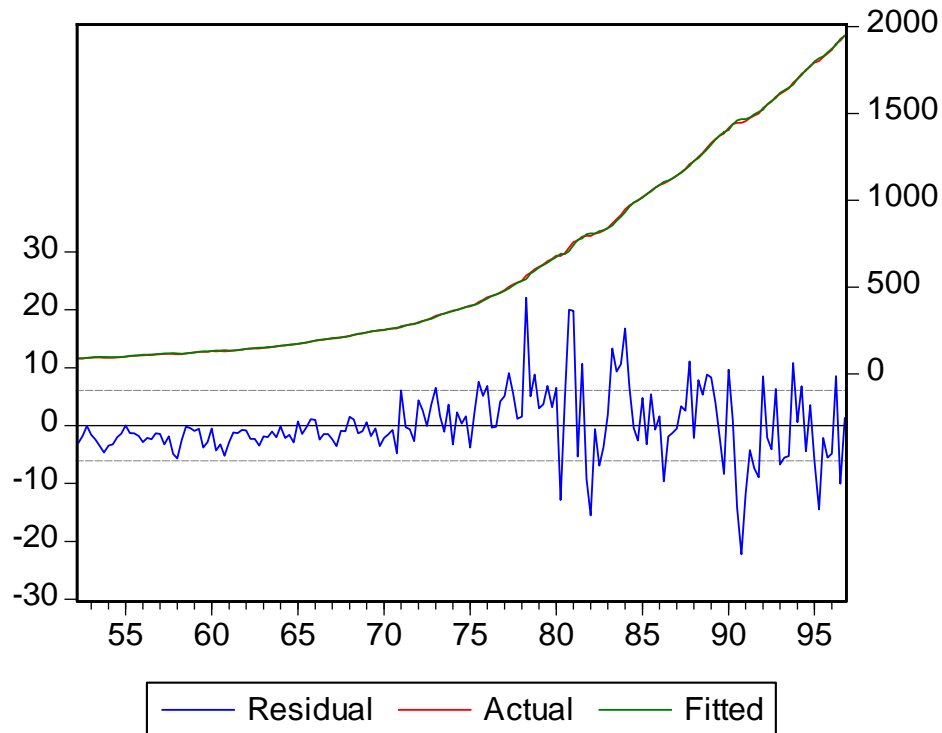
- Specify/Estimate...
- Forecast...
- Make Residual Series...
- Make Regressor Group
- Make Gradient Group
- Make Derivative Group
- Make Model
- Update Coefs from Equation

Phần dư từ phương trình hồi qui

Phần dư từ phương trình hồi qui được lưu giữ trong một đối tượng chuỗi có tên mặc định là RESID. RESID là một chuỗi thông thường chỉ được tạo ra do kết quả của phương trình hồi qui vừa mới được thực hiện. Cho nên, nếu muốn phân tích phần dư

¹ Representation

ta nên lưu phần dư với một tên gọi khác. Khi vừa ước lượng mô hình, nếu ta chọn **Resids** ở **Equation** ta sẽ có đồ thị hệ trục kép như sau:

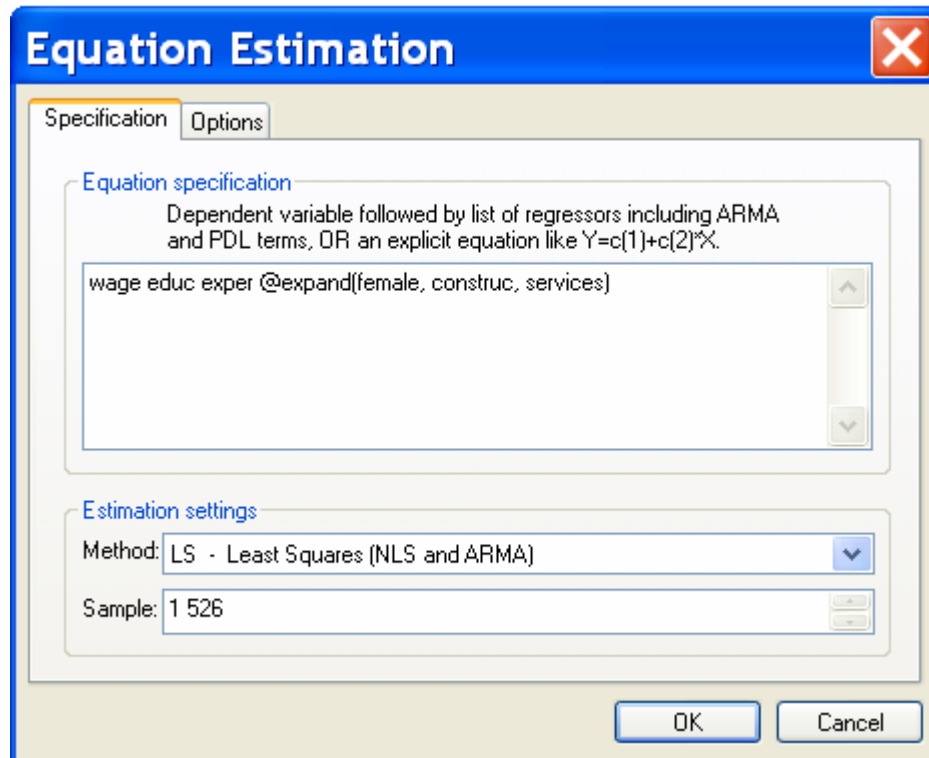


Biến giả trong Eviews

Để đưa biến giả vào mô hình hồi qui, thay vì phải tạo ra các biến này, Eviews đưa ra công thức hỗ trợ rất hữu ích như sau:

$$@EXPAND(D1, D2, \dots) \tag{2.23}$$

Ví dụ sử dụng **Chapter2.4.wf1** hồi qui biến wage theo các biến giáo dục, năm kinh nghiệm, giới thích, ngành xây dựng, và ngành dịch vụ như sau:



Kết quả ước lượng như sau:

Dependent Variable: WAGE
 Method: Least Squares
 Date: 02/16/08 Time: 11:13
 Sample: 1 526
 Included observations: 526

Variable	Coefficient	Std. Error	t-Statistic	Prob.
EDUC	0.606685	0.051740	11.72564	0.0000
EXPER	0.065625	0.010451	6.279253	0.0000
FEMALE=0,CONSTRUC=0,SERVICES=0	-1.737054	0.773135	-2.246767	0.0251
FEMALE=0,CONSTRUC=0,SERVICES=1	-3.396637	1.031271	-3.293641	0.0011
FEMALE=0,CONSTRUC=1,SERVICES=0	-1.078353	0.968035	-1.113960	0.2658
FEMALE=1,CONSTRUC=0,SERVICES=0	-3.869657	0.745000	-5.194174	0.0000
FEMALE=1,CONSTRUC=0,SERVICES=1	-4.569905	0.858127	-5.325444	0.0000
FEMALE=1,CONSTRUC=1,SERVICES=0	-3.839139	1.370391	-2.801492	0.0053
R-squared	0.319426	Mean dependent var	5.896103	
Adjusted R-squared	0.310229	S.D. dependent var	3.693086	
S.E. of regression	3.067198	Akaike info criterion	5.094498	
Sum squared resid	4873.190	Schwarz criterion	5.159370	
Log likelihood	-1331.853	Durbin-Watson stat	1.819378	

Một số kiểm định hồi qui cơ bản

Mỗi thủ tục kiểm định được trình bày dưới đây liên quan đến một giả thiết không (H_0) nhất định. Trong mỗi kết quả kiểm định thường gồm có một hoặc một số thống kê kiểm định và giá trị xác suất, p-value (càng thấp thì khả năng bác bỏ H_0 càng cao). Tuy nhiên, để hiểu các kiểm định ta cần hiểu các phân phối xác suất cơ bản cũng như

nội dung của kiểm định. Cho nên, các kiểm định này sẽ được áp dụng ngay sau nội dung bài giảng liên quan ở các chương sau.

Các kiểm định hệ số

- **Kiểm định Wald**

Đây là một kiểm định rất quan trọng trong phân tích hồi qui bội bằng cách tính một thống kê kiểm định dựa trên hồi qui không bị ràng buộc. Thống kê Wald cho biết mức độ các ước lượng không bị ràng buộc thỏa mãn các ràng buộc như thế nào dưới giả thiết không. Nếu các ràng buộc thực sự là đúng, thì các ước lượng không bị ràng buộc sẽ thỏa mãn các ràng buộc.

Để thực hiện kiểm định Wald (về ràng buộc hệ số) trên Eviews ta xem ví dụ sau đây (tập tin **Chapter2.5.wf1**). Giả sử ta có hàm sản xuất Cobb-Dougle có dạng:

$$Y = AK^{\beta_2}L^{\beta_3}e^{u_i} \tag{2.24}$$

Trong đó:

Y là sản lượng, K là vốn cố định, L là lao động, β_2 và β_3 lần lượt là hệ số co giãn của sản lượng theo vốn và lao động, và $\beta_2 + \beta_3$ là tính (lợi thế) kinh tế theo/nhờ qui mô (return to scale). Theo lý thuyết kinh tế ta biết:

- Lợi thế kinh tế tăng theo qui mô khi $\beta_2 + \beta_3 > 1$
- Lợi thế kinh tế không đổi theo qui mô khi $\beta_2 + \beta_3 = 1$
- Lợi thế kinh tế giảm theo qui mô khi $\beta_2 + \beta_3 < 1$

Nếu lấy log tự nhiên hai vế của phương trình (2.24) ta có:

$$\log(Y) = \log(A) + \beta_2\log(K) + \beta_3\log(L) + u_i \tag{2.25}$$

Đặt $\beta_1 = \log(A)$ và $A = e^{\beta_1}$, vậy ta có phương trình tương đương sau đây:

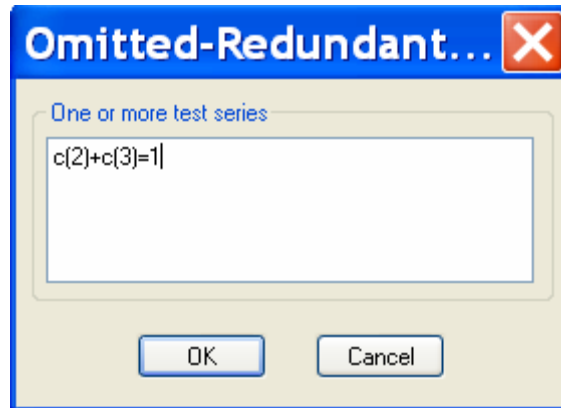
$$\log(Y) = \beta_1 + \beta_2\log(K) + \beta_3\log(L) + u_i \tag{2.26}$$

Kết quả ước lượng phương trình (2.25) như trong bảng sau:

Dependent Variable: LOG(Y)
 Method: Least Squares
 Date: 02/15/08 Time: 22:12
 Sample: 1955 1974
 Included observations: 20

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-1.652419	0.606198	-2.725873	0.0144
LOG(K)	0.845997	0.093352	9.062488	0.0000
LOG(L)	0.339732	0.185692	1.829548	0.0849
R-squared	0.995080	Mean dependent var		12.22605
Adjusted R-squared	0.994501	S.D. dependent var		0.381497
S.E. of regression	0.028289	Akaike info criterion		-4.155221
Sum squared resid	0.013604	Schwarz criterion		-4.005861
Log likelihood	44.55221	F-statistic		1719.231
Durbin-Watson stat	0.425667	Prob(F-statistic)		0.000000

Tổng các hệ số hồi qui của $\log(K)$ và $\log(L)$ dường như lớn hơn 1, nhưng để có kết luận tin cậy ta cần kiểm định giả thiết $H_0: \beta_2 + \beta_3 = 1$. Để thực hiện kiểm định Wald ta chọn **View/Coefficient Tests/Wald – Coefficient Restrictions ...** và nhập điều kiện ràng buộc vào hộp thoại soạn thảo như sau:



Lưu ý, nếu có nhiều ràng buộc khác nhau, thì mỗi ràng buộc cách nhau bằng một dấu phẩy. Eviews sẽ cho kết quả kiểm định như sau:

Wald Test:
Equation: Untitled

Test Statistic	Value	df	Probability
F-statistic	3.779440	(1, 17)	0.0686
Chi-square	3.779440	1	0.0519

Null Hypothesis Summary:

Normalized Restriction (= 0)	Value	Std. Err.
-1 + C(2) + C(3)	0.185730	0.095536

Restrictions are linear in coefficients.

Các giá trị thống kê sẽ được giải thích ở chương mô hình hồi qui bội. Ngoài ra, ta có thể đưa ra các điều kiện ràng buộc khác tùy vào phát biểu giả thiết.

Để quyết định bác bỏ hay chấp nhận H_0 , nếu là mô hình hồi qui tuyến tính ta so sánh giá trị F tính toán với giá trị F phê phán ở một mức ý nghĩa xác định. Ngược lại, nếu mô hình hồi qui phi tuyến ta so sánh giá trị chi bình phương tính toán với giá trị chi bình phương phê phán với số bậc tự do bằng số ràng buộc.

• Kiểm định bổ sót biến

Đây là một nội dung quan trọng trong kiểm định sai dạng mô hình. Ý tưởng của kiểm định này là khi ta đưa thêm biến vào mô hình và muốn biết các biến này có đóng góp có ý nghĩa vào việc giải thích sự thay đổi của biến phụ thuộc hay không. Giả thiết không của kiểm định này là các biến mới đưa thêm vào mô hình đồng thời không có ý nghĩa. Giả sử, với **Chapter2.3.xls**, lúc đầu ta chỉ ước lượng mô hình như sau:

$$\log(\text{GDP}_t) = B_1 + B_2 \log(\text{M1}_t) + B_3 \log(\text{RS}_t) + u_t \quad (2.27)$$

Hai điểm lưu ý với kiểm định này:

- Số quan sát trong hai mô hình phải bằng nhau.
- Áp dụng cho mọi phương pháp ước lượng miễn là phương trình hồi qui được xác định bằng cách liệt kê các biến chứ không phải bằng công thức.

Để thực hiện kiểm định bỏ sót biết ta chọn **View/Coefficient Tests/Omitted Variables – Likelihood Ratio** ... và nhập tên các biến nghi là bị bỏ sót cần được kiểm định (giả sử đó là TIME và PR) vào hộp thoại soạn thảo và được kết quả sau đây:

Omitted Variables: @TREND(1951Q4) PR

F-statistic	62.01102	Prob. F(2,175)	0.000000
Log likelihood ratio	96.43163	Prob. Chi-Square(2)	0.000000

Test Equation:
 Dependent Variable: LOG(GDP)
 Method: Least Squares
 Date: 02/16/08 Time: 00:14
 Sample: 1952Q1 1996Q4
 Included observations: 180

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.376707	0.310172	1.214511	0.2262
LOG(M1)	0.828530	0.068526	12.09079	0.0000
LOG(RS)	0.155011	0.010031	15.45254	0.0000
@TREND(1951Q4)	0.006215	0.000630	9.859122	0.0000
PR	-0.001373	0.095133	-0.014435	0.9885

R-squared	0.997850	Mean dependent var	5.991505
Adjusted R-squared	0.997800	S.D. dependent var	1.002533
S.E. of regression	0.047019	Akaike info criterion	-3.249158
Sum squared resid	0.386883	Schwarz criterion	-3.160464
Log likelihood	297.4242	F-statistic	20300.83
Durbin-Watson stat	0.211741	Prob(F-statistic)	0.000000

Để quyết định bác bỏ hay chấp nhận H_0 , nếu là mô hình hồi qui tuyến tính ta so sánh giá trị F tính toán với giá trị F phê phán ở một mức ý nghĩa xác định. Ngược lại, nếu mô hình hồi qui phi tuyến ta so sánh giá trị LR với giá trị chi bình phương phê phán với số bậc tự do bằng số ràng buộc.

• **Kiểm định thừa biến**

Đây cũng là một nội dung trong kiểm định sai dạng mô hình. Kiểm định này cho phép ta kiểm định xem một nhóm biến đưa vào mô hình có ý nghĩa thống kê hay không. Nói cách khác, đây là kiểm định xem các hệ số của một nhóm biến đưa vào mô hình có đồng thời bằng không hay không để quyết định có nên loại chúng ra khỏi mô hình hay không. Các điều kiện áp dụng kiểm định này cũng tương tự như kiểm định bỏ sót biến. Giả sử lúc đầu ta có mô hình như sau:

$$\log(\text{GDP}_t) = B_1 + B_2 \log(\text{M1}_t) + B_3 \log(\text{RS}_t) + B_4 \text{PR}_t + B_5 \text{TIME} + u_t \quad (2.28)$$

Để thực hiện kiểm định thừa biến ta chọn **View/Coefficient Tests/Redundant Variables – Likelihood Ratio ...** và nhập các biến cho rằng không cần thiết (ví dụ PR và TIME) vào hộp thoại soạn thảo và được kết quả sau:

F-statistic	62.01102	Prob. F(2,175)	0.000000
Log likelihood ratio	96.43163	Prob. Chi-Square(2)	0.000000

Test Equation:
 Dependent Variable: LOG(GDP)
 Method: Least Squares
 Date: 02/16/08 Time: 00:50
 Sample: 1952Q1 1996Q4
 Included observations: 180

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-1.448653	0.036617	-39.56272	0.0000
LOG(M1)	1.219905	0.007296	167.1909	0.0000
LOG(RS)	0.228400	0.009487	24.07574	0.0000
R-squared	0.996326	Mean dependent var	5.991505	
Adjusted R-squared	0.996284	S.D. dependent var	1.002533	
S.E. of regression	0.061113	Akaike info criterion	-2.735649	
Sum squared resid	0.661066	Schwarz criterion	-2.682433	
Log likelihood	249.2084	F-statistic	23996.62	
Durbin-Watson stat	0.282393	Prob(F-statistic)	0.000000	

Để quyết định bác bỏ hay chấp nhận H_0 , nếu là mô hình hồi qui tuyến tính ta so sánh giá trị F tính toán với giá trị F phê phán ở một mức ý nghĩa xác định. Ngược lại, nếu mô hình hồi qui phi tuyến ta so sánh giá trị chi bình phương tính toán với giá trị chi bình phương phê phán với bậc tự do bằng số ràng buộc.

Kiểm định phần dư

Eviews cung cấp các kiểm định các giả thiết rất quan trọng liên quan đến phần dư như phần dư có phân phối chuẩn, không có tương quan chuỗi, và phương sai đồng nhất.

• Kiểm định phân phối chuẩn của phần dư

Như ta sẽ biết ở chương 6 về phân tích hồi qui đơn, một giả định quan trọng trong các mô hình hồi qui tuyến tính cổ điển là các hạn nhiễu ngẫu nhiên trong mô hình hồi qui tổng thể (và vì thế phần dư - đại diện của hạn nhiễu ngẫu nhiên trong mô hình hồi qui mẫu) theo phân phối chuẩn.

Giả thiết H_0 : Phần dư của mô hình hồi qui có phân phối chuẩn

Phương pháp: Thống kê Jarque – Bera. Như đã trình bày thống kê JB có phân phối Chi bình phương với số bậc tự do là 2 ($\chi^2_{(2)}$).

Thực hiện trên Eviews: Từ kết quả ước lượng chọn **View/Residual tests/Histogram – Normality test**, rồi so sánh giá trị JB với giá trị Chi bình phương với số bậc tự do là 2.

• Biểu đồ tự tương quan - thống kê Q

Trong mô hình hồi qui tuyến tính cổ điển có giả định rằng các hạn nhiễu không có tương quan với nhau. Ngoài thống kê d Durbin-Watson, ta có thể sử dụng biểu đồ tự

tương quan và thống kê Q để kiểm định “chuỗi” phần dư của mô hình hồi qui có tương quan với nhau không. Biểu đồ tự tương quan đã được trình bày ở phần xử lý dữ liệu chuỗi. Để thực hiện kiểm định phần dư có tự tương quan hay không ta chọn **View/Residual Tests/Correlogram – Q Statistics ...**

- **Kiểm định nhân tử Lagrange**

Đây là một cách kiểm định khác với kiểm định Q để kiểm định tương quan chuỗi. Kiểm định này sẽ được trình bày ở chương 13 về lựa chọn dạng mô hình. Trên Eviews ta thực hiện kiểm định này bằng cách chọn **Views/Residual Tests/Serial Correlation LM Test ...**

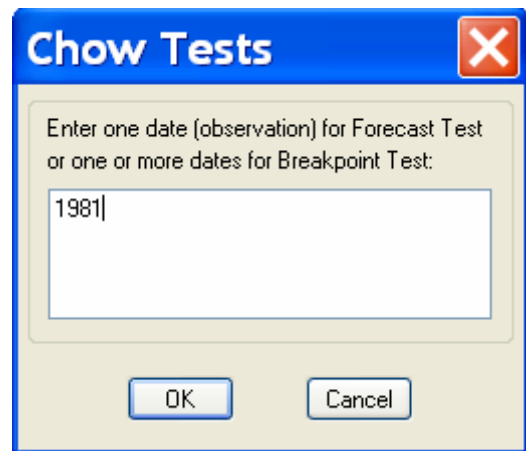
- **Kiểm định White về phương sai thay đổi**

Tương tự, mô hình hồi qui tuyến tính cổ điển cũng giả định các hạn nhiễu có phương sai đồng nhất. Để xem phương sai của nhiễu có đồng nhất hay không ta có thể sử dụng các kiểm định Park, kiểm định Glejser, kiểm định White, ... Nội dung các kiểm định này sẽ được trình bày ở chương 11 về phương sai thay đổi. Trên Eviews ta thực hiện kiểm định White bằng cách chọn hoặc **View/Residual Tests/White Heteroskedasticity (no cross terms)** hoặc **View/Residual Tests/White Heteroskedasticity (cross terms)**.

Kiểm định sự ổn định

- **Kiểm định Chow**

Mục đích của kiểm định Chow là xem liệu có sự thay đổi về mặt cấu trúc của mô hình hồi qui (đối với hồi qui chuỗi thời gian) giữa các giai đoạn khác nhau (do thay đổi chính sách hoặc cú sốc kinh tế) hay không. Nội dung của kiểm định này sẽ được trình bày ở chương 8 về phân tích hồi qui bội. Ta xét ví dụ trong **Chapter2.6.txt**. Sau khi ta hồi qui tiết kiệm theo thu nhập và thực hiện kiểm định như sau **View/Stability Tests/Chow Breakpoint Test ...** và ta nhập mốc thời gian vào hộp thoại soạn thảo để có kết quả như sau:



Chow Breakpoint Test: 1981

F-statistic	11.89335	Prob. F(2,22)	0.000315
Log likelihood ratio	19.05673	Prob. Chi-Square(2)	0.000073

- **Kiểm định RESET của Ramsay**

Mục đích của kiểm định này là xem có bỏ sót biến quan trọng trong mô hình hồi qui hay không (nhất là khi không có số liệu về biến bỏ sót đó). Nội dung của kiểm định này sẽ được trình bày ở chương 8 về lựa chọn dạng mô hình. Sau khi ước lượng, để kiểm định xem liệu mô hình có sót biến hay không ta chọn **View/Stability Tests/Ramsay RESET Test ...**

|
TRƯỜNG ĐẠI HỌC TÔN ĐỨC THẮNG
KHOA KHOA HỌC XÃ HỘI VÀ NHÂN VĂN
----- ✍️ 📖 ✍️ -----



MÔN HỌC

XỬ LÝ VÀ PHÂN TÍCH DỮ LIỆU NGHIÊN CỨU VỚI SPSS

GIẢNG VIÊN: CN. TẠ XUÂN HOÀI

Xử lý và phân tích dữ liệu nghiên cứu với SPSS

Đối tượng sử dụng:
tất cả sinh viên ứng dụng thống kê trong nghiên cứu

Đại học Tôn Đức Thắng
Khoa Khoa học xã hội và Nhân văn

Mục đích môn học

SPSS (Statistical Package for Social Sciences) là một phần mềm quản lý cơ sở dữ liệu và cho phép thực hiện nhiều phép phân tích thống kê chuyên nghiệp, đa năng, dùng:

- Trình bày dữ liệu: bảng tần suất, phần trăm, trị trung bình, độ lệch chuẩn, trình bày biểu đồ
- Kiểm định thống kê: xem xét mối liên hệ, sự tương quan của dữ liệu

SPSS có khả năng xử lý và phân tích các dữ kiện nghiên cứu có dung lượng mẫu từ vài chục cho đến vài chục ngàn mẫu hay nhiều hơn nữa tùy theo khả năng của bộ nhớ máy vi tính.

Cấu trúc môn học:

- A. Nguyên tắc tạo khuôn nhập liệu
 - B. Thống kê mô tả (Trình bày dữ liệu)
 - C. Thống kê suy diễn (Kiểm định dữ liệu)
-

Tài liệu tham khảo:

1. Thống kê ứng dụng trong kinh tế xã hội. Hoàng Trọng (2007)
2. Phân tích dữ liệu nghiên cứu với SPSS. Hoàng Trọng (2005)

Thông tin

Download tập tin **data thuchanh.sav** để thực hành

Địa chỉ: - taxuanhoai@gmail.com

- phantichdulieu@yahoo.com.vn

Xử lý và phân tích dữ liệu nghiên cứu với SPSS

Chương 1:

**Phân loại dữ liệu, mã hóa
và tạo khuôn nhập liệu**

Các giai đoạn của tiến trình nghiên cứu

Chuẩn bị

Đề cương chi tiết

Bộ công cụ thu thập thông tin

**Tiến hành
thu thập dữ liệu**

Dữ liệu thô

Báo cáo tổng hợp

**Xử lý, phân tích
thông tin**

Vị trí của kỹ thuật xử lý và phân tích dữ liệu



Sơ đồ tư duy thống kê

Mức độ chính xác
của mô hình thống kê

Tri thức

Sự kiện

Thông tin

Dữ liệu

Mức độ cải thiện các quyết định

1. Phân loại dữ liệu

+ **Dữ liệu định tính:** Là loại dữ liệu phản ánh tính chất, thuộc tính của đối tượng nghiên cứu. Chỉ mô tả sự hơn kém không tính được trị trung bình.

Thu thập bằng thang đo định danh và thang đo thứ bậc.

Ví dụ: Dữ liệu về

- . thành phần dân tộc
- . giới tính
- . màu sắc
- . Loại hình phương tiện giao thông
- ...

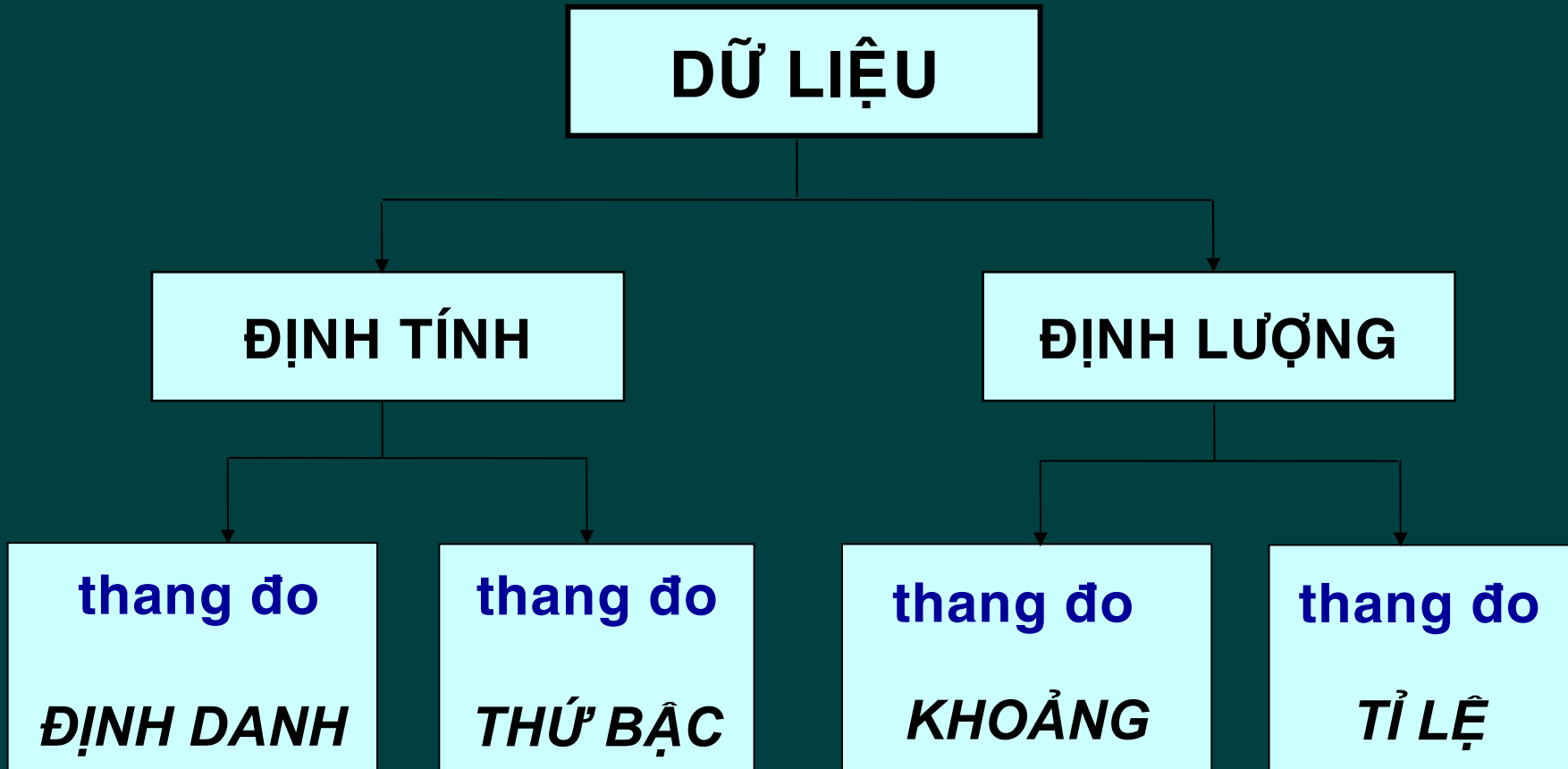
1. Phân loại dữ liệu

+ **Dữ liệu định lượng:** Là loại dữ liệu phản ánh mức độ hay mức độ hơn kém và tính được trị trung bình. Thu thập bằng thang đo khoảng cách và thang đo tỉ lệ.

Ví dụ: Dữ liệu về

- . tuổi
- . thu nhập trung bình tháng
- . số sinh viên vắng mặt
- . trọng lượng cơ thể
- ...

SƠ ĐỒ PHÂN LOẠI DỮ LIỆU



Thang đo

+ **Danh nghĩa (Nominal):** Là loại thang đo dùng cho các đặc điểm thuộc tính, dùng để phân loại đối tượng. Khi thống kê người ta thường sử dụng các mã số để qui ước, giữa các con số này không có quan hệ hơn kém và không ý nghĩa toán học.

Ví dụ: Tình trạng hôn nhân của anh/chị hiện nay:

- 1. Độc thân*
- 2. Có gia đình*
- 3. Ly dị*
- 4. Trường hợp khác*

Thang đo

+ **Thứ bậc (Ordinal)**: Là loại thang đo dùng cho các đặc điểm thuộc tính, các giá trị được sắp xếp theo trật tự tăng hoặc giảm dần và có mối quan hệ thứ bậc hơn kém.

Thực chất thang đo thứ bậc là thang đo định danh các giá trị được sắp xếp theo thứ bậc.

Ví dụ: Học lực

- 1. Yếu, kém*
- 2. Trung bình*
- 3. Khá*
- 4. Giỏi*
- 5. Xuất sắc*

Thang đo

+ **Khoảng cách (Interval):** Là loại thang đo dùng cho các đặc điểm số lượng, là thang đo thứ bậc có các khoảng cách đều nhau và liên tục. Dãy số này có hai cực ở hai đầu dãy số thể hiện hai trạng thái đối nghịch nhau.

Dữ liệu tính toán cộng trừ có ý nghĩa

Ví dụ: Vấn đề quan tâm (không quan tâm: 1 ... quan tâm nhất: 5)

	Không quan tâm			Rất quan tâm	
Đạo đức	1	2	3	4	5
Trình độ văn hóa	1	2	3	4	5
Trình độ chuyên môn	1	2	3	4	5

Thang đo

+ **Tỉ lệ (Ratio):** Là loại thang đo dùng cho đặc tính số lượng. Thang đo tỉ lệ có đầy đủ đặc tính của thang đo khoảng cách. Ngoài ra nó còn có một trị số “0” thật, cho phép lấy tỉ lệ so sánh giữa hai giá trị của biến số

Dữ liệu tính toán tất cả đều có ý nghĩa.

Ví dụ:

- *Tuổi của anh/chị: tuổi*
- *Anh/chị đã vay ngân hàng bao nhiêu tiền:
. VNĐ*

Thang đo

Chú ý:

Sự khác nhau giữa thang đo khoảng và thang đo tỉ lệ:

- . Điểm 0 trong thang đo tỉ lệ là một trị số thật
- . Trong thang đo khoảng sự so sánh về mặt tỉ lệ giữa các giá trị không có ý nghĩa

Trong xử lý SPSS thường gộp chung thang đo khoảng cách và thang đo tỉ lệ thành thang đo định lượng (Scale). → định danh (Nominal); thứ bậc (Ordinal) và định lượng (Scale).

2. Giới thiệu SPSS for Windows - làm việc với SPSS

2.1. Khởi động SPSS – Kết thúc SPSS

+ Khởi động:

- Menu *Programs* ▪ *SPSS for Windows 11.5* màn hình SPSS sẽ xuất hiện.

+ Kết thúc SPSS:

- Menu *File* ▪ *Exit*

- Sử dụng tổ hợp phím: *Alt + F4*

- Nhấp nút trái trở vào biểu tượng chữ thập trắng nền đỏ góc phải trên màn hình.

2. Giới thiệu SPSS for Windows - làm việc với SPSS

2.2. Thực đơn, công cụ và một số thuật ngữ trong SPSS

+ Nội dung chủ yếu của Menu:

File: khởi tạo file mới, đóng mở, lưu file, in ấn, thoát

Edit: undo, cắt dán, chọn, tìm kiếm, thay thế, xác lập các mặc định (options)

View: cho hiện dòng trạng thái, thanh công cụ, chọn font chữ

Data: định nghĩa biến, thêm biến, đi đến quan sát, xếp thứ tự, ghép file, chia file chọn quan sát.

2. Giới thiệu SPSS for Windows - làm việc với SPSS

Transform: tính toán, mã hóa lại các biến ...

Analyze: thực hiện các thủ tục thống kê như: tóm tắt dữ liệu, lập bảng tổng hợp, so sánh trung bình của hai đám đông, phân tích phương sai, tương quan và hồi quy, và các phương pháp phân tích đa biến.

Graphs: tạo các biểu đồ và đồ thị

Utilities: tìm hiểu thông tin về các biến, file

Windows: sắp xếp các cửa sổ làm việc trong SPSS, di chuyển giữa các cửa sổ làm việc

2. Giới thiệu SPSS for Windows - làm việc với SPSS

+ Vùng nội dung:

Variable View: là màn hình khuôn nhập liệu

Data View: là màn hình dữ liệu

Để chuyển đổi hai cửa sổ làm việc trên, chỉ cần nhấp vào Data View hay Variable View nằm ở góc trái màn hình.

2. Giới thiệu SPSS for Windows - làm việc với SPSS

2.3. Truy xuất một tập tin dữ liệu (Data) – một kết quả (Output) có sẵn

- (1) *Menu* ▪ *File* ▪ *Open* ▪ *Data (dữ liệu)/Output (kết quả)...* cửa sổ Open Data/Open Output xuất hiện.
- (2) Nhấp trên biểu tượng chọn ổ đĩa cần đến, sau đó nhấp đúp nút trái con trỏ chọn tập tin Data/Output cần chọn, tập tin cần dùng xuất hiện.

2. Giới thiệu SPSS for Windows - làm việc với SPSS

2.4. Các thuộc tính của Biến:

Name: tên biến

Type: kiểu biến

Decimals: số thập phân

Label: nhãn biến

Values: giá trị biến (mã hóa)

Missing: giá trị khuyết

Measure: thang đo (loại dữ liệu)

3. Nguyên tắc tạo khuôn nhập liệu

+ Nguyên tắc tạo khuôn:

- **Name:** 8 ký tự, không có ký tự đặc biệt, không có khoảng trắng, không bắt đầu bằng chữ số và không đặt trùng tên.
- **Type:** Numeric (kiểu số) dùng cho biến định lượng và biến định tính đã được mã hóa; String (kiểu chuỗi) dùng cho biến định tính không có mã hóa.
- **Label:** 255 ký tự, dùng để định nghĩa biến.
- **Values:** Chỉ mã hoá thang đo định tính, không cần phải mã hoá thang đo định lượng.
- **Missing:** nhập giá trị mà biến không xử lý.
- **Measure:** thang đo định danh: *nominal*; thang đo thứ bậc: *ordinal* ; thang khoảng và thang đo tỉ lệ: *scale*

3. Nguyên tắc tạo khuôn nhập liệu

+ Nguyên tắc tạo biến (Variables View):

Một câu hỏi có thể tạo ra một biến và cũng có thể tạo ra nhiều biến.

Chú ý câu hỏi nhiều lựa chọn (Multiple Answer - MA), tạo ra nhiều biến. Có hai cách tạo biến:

- . Kiểu Dichotomies: biến có hai trạng thái (có/không; yêu/ghét; đồng ý/không đồng ý)
- . Kiểu Categories: biến có nhiều trạng thái

3. Nguyên tắc tạo khuôn nhập liệu

+ Nguyên tắc nhập liệu (Data View):

- Thông tin của mỗi đối tượng trả lời tương ứng với 1 dòng (1 case).
- Mỗi loại thông tin thu thập tương ứng với 1 cột (1 variable).
- Nhập liệu từ trái qua phải theo từng dòng; xong 1 phiếu (1 dòng) thì chuyển sang phiếu khác (sang dòng mới).

4. Thay đổi mặc định của chương trình

- + Hiển thị tên hoặc nhãn biến trong ô chứa biến nguồn khi thực hiện các thủ tục xử lý:
 - Menu ▪ Edit ▪ Options hộp thoại xuất hiện
 - Chọn General ▪ Display Name (hiển thị tên) hoặc Display Label (hiển thị nhãn) ▪ Apply ▪ OK.
- + Điều chỉnh độ rộng và số lẻ biến kiểu Numeric:
 - Menu ▪ Edit ▪ Options hộp thoại xuất hiện
 - Chọn Data ▪ With (thay đổi độ rộng) hoặc Decimal (số thập phân) ▪ Apply ▪ OK.

4. Thay đổi mặc định của chương trình

+ Hiển thị số hoặc chữ trong Data View:

- Menu ▫ View ▫ Value Label (hiển thị chữ) hoặc không chọn Value Label (hiển thị số) ▫ Apply ▫ OK.

+ Hiển thị tiếng Việt trong Data View:

- Menu ▫ View ▫ Font (hệ chữ); Font Style (kiểu chữ) và Size (Mức độ chữ) ▫ Apply ▫ OK.

+ Hiển thị tiếng Việt trong Output:

- Menu ▫ Edit ▫ Options hộp thoại xuất hiện
- Chọn Pivot Tables ▫ Tablelook
- Chọn tập tin tiếng Việt Boxed VNI Helve condense.tlo
 - Set Tablelook Directory ▫ Apply ▫ OK.

Bài tập:

1. Cài chương trình SPSS 13.0 vào máy
2. Tạo khuôn nhập liệu cho một bảng hỏi mẫu

Xử lý và phân tích dữ liệu nghiên cứu với SPSS

Chương 2:

Trình bày dữ liệu
(Thống kê mô tả)

1. Lập bảng tần số

- + **Mục đích:** dùng bảng tần số để thống kê mẫu nghiên cứu theo từng biến và lọc dữ liệu.
- + **Đối tượng:** có thể được thực hiện với các biến số (kiểu định tính và định lượng). Trong trường hợp biến định lượng liên tục có quá nhiều giá trị
- + **Thao tác:** Menu **▪ Analyze ▪ Descriptive Statistics ▪ Frequencies** hộp thoại xuất hiện **▪ Chọn biến cần lập bảng tần số trong ô biến nguồn đưa vào ô Variables ▪ OK.** Kết quả bảng tần số xuất hiện.

1. Lập bảng tần số

+ Các thông số trong bảng Statistics:

* Dòng Valid: số quan sát hợp lệ (số người trả lời)

* Dòng Missing: số quan sát bị thiếu dữ liệu (số người không trả lời)

+ Các thông số trong bảng Frequencies:





* Cột Frequencies: tần số

* Cột Percent: tần suất/phần trăm

* Cột Valid Percent: phần trăm hợp lệ

* Cột Cumulative Percent: Phần trăm cộng dồn các phần trăm hợp lệ

2. Tính toán các đại lượng thống kê mô tả

- + **Mục đích:** Tính các đại lượng thống kê cho từng biến.
- + **Đối tượng:** Chỉ được tính cho các biến định lượng. Nếu dùng các đại lượng thống kê mô tả đối với các biến định tính, thì kết quả sẽ không có ý nghĩa.
- + **Thao tác:** Menu  Analyze  Descriptive Statistics  Descriptive  hộp thoại xuất hiện
Chọn biến định lượng cần tính trong ô biến nguồn đưa vào ô Variables

2. Tính toán các đại lượng thống kê mô tả

- Chọn các đại lượng thống kê cần tính trong **Options**:
 - . Mean: trung bình cộng . Sum: tổng cộng
 - . Std. Deviation: độ lệch chuẩn . SE. mean: sai số chuẩn
 - . Range: khoảng giữa giá trị cao nhất và giá trị thấp nhất
 - . Minimum: giá trị nhỏ nhất . Maximum: giá trị lớn nhất
- Chọn cách sắp xếp thứ tự kết quả trong **Options**:
 - . Ascending mean: trật tự tăng dần
 - . Descending mean: trật tự giảm dần
- Chọn **Continue** ▪ **OK**. Kết quả xuất hiện

3. Lập bảng tổng hợp nhiều biến (bảng chéo)

- + **Mục đích:** thống kê mẫu theo nhiều biến và dùng làm sạch dữ liệu.
- + **Đối tượng:** kết hợp các biến định tính và biến định lượng với nhau.
- + **Các dạng bảng chéo:**
 - Bảng 1 định tính – 1 định tính
 - Bảng 1 định tính – 2 định tính (hay nhiều định tính)
 - Bảng 1 định lượng – 1 định tính
 - Bảng 1 định lượng – 2 định tính (hay nhiều định tính)

(chú ý: trong một bảng không thể xử lý 2 biến định lượng)

3.1. Lập bảng 1 biến định tính – 1 biến định tính

+ Kiểu bảng Basic:

- Từ Menu ▫ Analyze ▫ Tables ▫ Basic Tables hộp thoại xuất hiện:

- . Down: ô chứa biến dòng
- . Across: ô chứa biến cột
- . Statistics: chọn hàm tính các đại lượng thống kê
- . Layout: sắp xếp các dữ liệu trong bảng số liệu
- . Total: tính tổng dòng và tổng cột

Lưu ý: Thông thường biến phụ thuộc đặt trong ô dòng và biến biến độc lập đặt trong ô cột. Nhưng thực tế, biến ít giá trị được đặt trong ô cột.

3.1. Lập bảng 1 biến định tính – 1 biến định tính

- Chọn hàm thống kê trong Statistics. Với biến định tính các hàm thường dùng là:
 - . Count: tần số
 - . Row%: phần trăm theo dòng
 - . Col%: phần trăm theo cột
- Sắp xếp số liệu trong bảng dùng hộp thoại Layout:
 - . Across the top: sắp xếp theo cột (chọn hàm Col%)
 - . Down the left side: sắp xếp theo dòng (chọn hàm Row%)
 - . In separate table: sắp xếp theo các bảng riêng, theo từng hàm thống kê.

3.1. Lập bảng 1 biến định tính – 1 biến định tính

+ Kiểu bảng General:

- Từ Menu **Analyze** **Table** **General Tables** hộp thoại xuất hiện:
 - . Rows: ô chứa biến dòng
 - . Columns: ô chứa biến cột
 - . Edit Statistics: chọn hàm thống kê và điều chỉnh định dạng số liệu tính ra
 - . Statistics Labels Appear: sắp xếp các đại lượng thống kê tính ra
 - . Omit Label: bỏ nhãn biến đang chọn
 - . Total: tính tổng dòng, tổng cột

3.2. Lập bảng 1 biến định tính – 2 biến định tính

+ Kiểu bảng Basic:

Thao tác giống lập bảng 2 biến định tính, nhưng chú ý:

- . Across: chứa hai biến định tính
- . All Combination: hai biến trong cùng ô lồng ghép nhau.
- . Each Separately: hai biến độc lập không cùng ô.

+ Kiểu bảng General:

Thao tác giống lập bảng 2 biến định tính, nhưng chú ý:

- . Statistics Labels Appear: kết hợp hoặc tách thành các bảng rời nhau
- . UnNets: hai biến cột không ngang cấp, chọn biến lùi vào trong
- . Omit Label: bỏ nhãn biến khi cần thiết

3.3. Lập bảng 1 biến định lượng – 1 biến định tính

Thao tác:

- Menu ▫ Analyze ▫ Tables ▫ Basic Tables hộp thoại xuất hiện:
 - . Down hoặc Across: dùng đưa biến định tính vào xử lý.
 - . Summaries: ô chứa biến định lượng cần tính (có thể đưa nhiều biến định lượng có liên quan vào để xử lý cùng một lúc, nếu cần thiết)
 - . Statistics: chọn hàm thống kê.
- Chọn Continue ▫ OK. Kết quả xuất hiện

3.4. Lập bảng 1 biến định lượng – 2 biến định tính

Thao tác:

- Từ Menu **Analyze** **Tables** **Basic Tables** hộp thoại xuất hiện.
- Thao tác như bảng 1 biến định lượng – 1 biến định tính, nhưng cần xác định biến định tính nào đưa vào dòng (Across) và biến định tính nào đưa vào cột (Down).

4. Họạ đồ

Pie (Biểu đồ tròn): sử dụng đối với dữ liệu một biến định tính dưới dạng tần số hay %, có ít nhóm thông tin và có sự so sánh hơn.

Bar (Biểu đồ thanh ngang): sử dụng đối với dữ liệu một hoặc nhiều biến định tính dưới dạng tần số hay %, khi có rất nhiều nhóm thông tin.

Column (Biểu đồ cột): sử dụng đối với dữ liệu một hoặc nhiều biến định tính dưới dạng tần số hay %, khi có ít nhóm thông tin.

Line (Đồ thị đường gấp khúc): sử dụng đối với dữ liệu định lượng có sự biến thiên theo thời gian.

Area (Đồ thị diện tích): sử dụng đối với dữ liệu định lượng, dùng để nhấn mạnh sự thay đổi theo thời gian.

4. Họa đồ

Thao tác:

- Menu ▫ Graphs ▫ chọn một trong các dạng:
 - . Pie
 - . Bar
 - . Column
 - . Line
 - . Area

Nguyên tắc chung để thực hiện biểu đồ, đồ thị:

- + Để vẽ biểu đồ, đồ thị thường sử dụng Excel
- + Chuyển kết quả xử lý từ SPSS sang Excel
- + Dùng các lệnh vẽ đồ thị trong Excel để thực hiện

5. Xử lý câu hỏi Multiple Answer (MA)

Xử lý câu hỏi MA cần ghép các biến tạo bởi câu hỏi này lại với nhau thành một biến đại diện. Sau đó xử lý trên biến đại diện.

+ **Tạo biến đại diện** (ghép câu hỏi MA):

- Menu ▫ Analyze ▫ Tables ▫ General Tables ▫ hộp thoại xuất hiện
- Chọn Mult Response Set (lệnh ghép các biến trong cùng một câu hỏi MA) hộp thoại ghép biến xuất hiện:

Lưu ý :

- *Các biến đại diện đã tạo không tồn tại trong tập tin dữ liệu nguồn (data).*
- *Các biến đại diện chỉ khai báo một lần và sử dụng mãi cho tất cả các lần xử lý.*

5. Xử lý câu hỏi Multiple Answer (MA)

- Set Definition: ô chứa biến nguồn
- Variables in Set: ô đưa các biến trong cùng một câu hỏi MA vào để ghép biến
- Variables Are Coded As: khai báo cách mã hoá biến các câu hỏi MA
 - . Dichotomies: biến có hai trạng thái
 - . Categories: biến có nhiều trạng thái
- Name: đặt tên biến đại diện
- Label: đặt nhãn biến đại diện
- Add: xác nhận biến đại diện
- Mult Response Set: ô cập nhật biến đại diện vào danh sách
- **Save**, trở về hộp thoại ban đầu

5. Xử lý câu hỏi Multiple Answer (MA)

+ Lập bảng tổng hợp với câu hỏi MA:

- Menu Analyze Tables General Tables hộp thoại xuất hiện
 - Reset, chọn các biến cần xử lý chuyển vào ô tương ứng, luôn cả biến đại diện.
- Rows: ô tạo dòng
- Columns: ô tạo cột
- Layers: tạo bảng kết quả tách biệt
 - OK: kết thúc

Chọn hàm thống kê, cách hiển thị hàm thống kê và bỏ nhấn biến đang chọn, thao tác giống như lập bảng General.

6. Một số xử lý trên biến

+ Tính toán trên biến (Compute):

- Mục đích: Dùng để chuyển đổi dữ liệu định lượng, dựa trên những thông tin chứa trong một vài biến nguồn.
- Ví dụ: Tạo biến năm sinh từ thông tin về tuổi; tạo biến chứa tỉ lệ người xem quảng cáo trên số người đọc báo SGTT.
- Thao tác: từ Menu ▫ Transform ▫ Compute hộp thoại xuất hiện:
 - Target Variable: đặt tên (Name) cho biến mới
 - Numeric Expression: nhập biểu thức tính
 - OK, biến mới được tạo ra
 - Khai báo nhãn (Label) và kiểu biến (Type) cho biến mới trong Variables View

6. Một số xử lý trên biến

+ Mã hóa lại biến (Recode):

- Mục đích: Khi các biến có quá nhiều giá trị (nhất là biến định lượng), bảng tần số ít có ý nghĩa trong việc tóm tắt và trình bày. Cần phải mã hóa lại để chỉ còn ít nhóm hoặc ít giá trị dễ dàng trong việc phân tích và trình bày rõ ràng hơn.

- Ví dụ: Tạo biến độ tuổi từ biến tuổi, từ biến thu nhập cá nhân (dữ liệu định lượng) chuyển sang biến nhóm thu nhập cá nhân (dữ liệu định tính).

- Thao tác: Từ Menu **Transform** **Recode** **Into Different Variable** hộp thoại xuất hiện:

6. Một số xử lý trên biến

- Numeric Variable -> Output Variable: đưa biến cần mã hóa lại vào
 - Name: đặt tên biến mới
 - Label: đặt nhãn cho biến mới
 - Change
 - Old And New Value hộp thoại xuất hiện:
 - . Value: từng giá trị cũ rời rạc ứng với một giá trị mới
 - . Range: một khoảng giá trị cũ ứng với một giá trị mới
 - . Range Lowest through: khoảng giá trị cũ thấp nhất
 - . Range through highest: khoảng giá trị cũ cao nhất
 - . New Value: giá trị mới tương ứng
 - . Add: xác định xong một cặp giá trị vào ô Old -> New
 - Continue
 - OK, biến mới được tạo ra
 - Khai báo nhãn (Label) và kiểu biến (Type) cho biến mới

Bài tập

1. Truy xuất tập tin Data thuchanh.sav
2. Chọn biến và trình bày dữ liệu phù hợp với từng công cụ trình bày dữ liệu cơ bản.

Xử lý và phân tích dữ liệu nghiên cứu với SPSS

Chương 3:

Kiểm định thống kê
(Thống kê suy diễn)

1. Kiểm định tính độc lập (Chi – Square)

+ Mục đích và vận dụng:

Kiểm định Chi – Square được sử dụng để kiểm định xem có tồn tại mối quan hệ giữa hai yếu tố đang nghiên cứu trong tổng thể hay không.

+ Đối tượng kiểm định:

Sử dụng kiểm định Chi – Square khi hai yếu tố đang nghiên cứu là biến định tính hay biến định lượng rời rạc có ít giá trị:

- . Biến phụ thuộc: dữ liệu định tính
- . Biến độc lập: dữ liệu định tính

1. Kiểm định tính độc lập (Chi – Square)

+ Cơ sở lý thuyết:

Giả thuyết:

- . H_0 : hai biến nghiên cứu độc lập (không liên hệ)
- . H_1 : hai biến có liên hệ với nhau

(1) Để kiểm định giả thuyết H_0 trước hết cần tính các tần số lý thuyết E_{ij} tại từng ô trong bảng phân tổ kết hợp trên cơ sở giả thuyết, rằng giả thuyết H_0 đúng. Các tần số lý thuyết này được tính theo công thức sau:

$$E_{ij} = \frac{R_i \cdot C_j}{n}$$

R_i : tổng số quan sát của dòng thứ i

C_j : tổng số quan sát của cột thứ j

1. Kiểm định tính độc lập (Chi – Square)

(2) Sau đó tính chênh lệch giữa tần số thực tế với tần số lý thuyết trong từng ô, bình phương lên rồi chia cho tần số lý thuyết. Cộng tất cả các kết quả tính toán ở từng ô này lại với nhau, thể hiện bằng công thức:

Đại lượng dùng để kiểm định là:

$$\chi^2 = \sum_{i=1}^r \sum_{j=1}^c \frac{(O_{ij} - E_{ij})^2}{E_{ij}}$$

Trong đó:

χ^2 : đại lượng Chi - Squares dùng để kiểm định

O_{ij} : tần số quan sát thực tế trong các ô của bảng chéo

E_{ij} : tần số quan sát lý thuyết trong các ô của bảng chéo

c : số cột của bảng

r : số dòng của bảng

1. Kiểm định tính độc lập (Chi – Square)

$\chi^2_{(r-1)(c-1),\alpha}$ là trị số tra bảng phân phối Chi - Square.

Với số bậc tự do (df) bằng $(r-1) \times (c-1)$ và mức ý nghĩa α .
Tiêu chuẩn quyết định là:

+ Bác bỏ **H₀** nếu: $\chi^2 > \chi^2_{(r-1)(c-1),\alpha}$ □ Sig. < $\alpha = 0,05$

+ Chấp nhận **H₁** nếu : $\chi^2 \leq \chi^2_{(r-1)(c-1),\alpha}$ ▪ Sig < $\alpha = 0,05$

1. Kiểm định tính độc lập (Chi – Square)

+ Thao tác kiểm định:

- Từ Menu **Analyze** ▪ **Descriptive Statistics** ▪ **Crosstabs** hộp thoại xuất hiện:
 - . Row: ô đưa biến phụ thuộc vào
 - . Column: ô đưa biến tác động vào
 - . Statistics: chọn công cụ kiểm định Chi - Square
 - . Cell: chọn hàm, chọn cách thể hiện hàm thống kê
- Sau khi chọn xong nhấn OK, kết quả kiểm định Chi – Square xuất hiện.

1. Kiểm định tính độc lập (Chi – Square)

Lưu ý:

- Các ô có tần số lý thuyết nhỏ hơn 5 chiếm không lớn hơn tỉ lệ 20% trong tổng số ô trong bảng, việc kiểm định Chi bình phương sẽ mang lại kết quả chấp nhận được.
- Kiểm định Chi – Square dùng kiểm định xem có tồn tại mối quan hệ giữa hai yếu tố định tính với thang đo định danh hoặc thang đo thứ bậc. Nhưng nếu hai yếu tố định tính với thang đo thứ bậc thì cần sử dụng thêm các đại lượng kiểm định: **Kendall's tau-b**; **Kendall's tau-c**; **Somers'd** hoặc **Gamma**, các đại lượng này giúp phát hiện ra mối liên hệ tốt hơn Chi - Square.

2. Kiểm định trung bình (t - Test)

+ Mục đích và vận dụng:

Cần so sánh trị trung bình về một chỉ tiêu nghiên cứu nào đó giữa hai đối tượng ta quan tâm (Nghĩa là so sánh trung bình giữa hai giá trị)

+ Đối tượng kiểm định:

- . Biến phụ thuộc: dữ liệu định lượng (tính trung bình)
- . Biến độc lập: dữ liệu định tính (so sánh 2 giá trị)

+ Các dạng kiểm định Trung bình:

- . Kiểm định trong trường hợp hai mẫu độc lập
- . Kiểm định trong trường hợp hai mẫu phụ thuộc

2.1. Kiểm định trong trường hợp hai mẫu độc lập

+ Cơ sở lý thuyết:

(1) Giả sử các mẫu được chọn ngẫu nhiên độc lập từ hai tổng thể có:

- phân phối chuẩn: X, Y
- trung bình: μ_x, μ_y
- phương sai: σ_x^2, σ_y^2
- trung bình mẫu: \bar{X}, \bar{Y}

Ta cần kiểm định giả thuyết:

$$H_0: \mu_x - \mu_y = 0 \quad (0 \text{ cho trước})$$

$$H_1: \mu_x - \mu_y \neq 0$$

2.1. Kiểm định trong trường hợp hai mẫu độc lập

Dựa vào mẫu ta đưa ra quy tắc chấp nhận hay bác bỏ giả thuyết trên với mức ý nghĩa α và giá trị kiểm định:

$$t = \frac{\bar{x} - \bar{y}}{\sqrt{\frac{\sigma_x^2}{n_x} + \frac{\sigma_y^2}{n_y}}}$$

Biết $\alpha = 0,05$ (độ tin cậy 95%), ta tra bảng t – test (sách Lý thuyết thống kê) tìm được $t_{\alpha/2}$

+ Bác bỏ giả thuyết H_0 khi: $|t| > t_{\alpha/2}$

+ Chấp nhận giả thuyết H_0 khi: $|t| \leq t_{\alpha/2}$

2.1. Kiểm định trong trường hợp hai mẫu độc lập

Lưu ý: Trường hợp chưa biết phương sai tổng thể kích thước mẫu phải lớn hơn 30, ta vẫn có thể dùng công thức như trên và thay thế:

$$\sigma_x^2 = s_x^2, \sigma_y^2 = s_y^2$$

Công thức sẽ là:

$$t = \frac{\bar{x} - \bar{y}}{\sqrt{\frac{S_x^2}{n_x} + \frac{S_y^2}{n_y}}}$$

\bar{x} : là trung bình mẫu nhóm x

\bar{y} : là trung bình mẫu nhóm y

S_x^2 : là phương sai mẫu nhóm x

S_y^2 : là phương sai mẫu nhóm y

n_x : là cỡ mẫu của nhóm x

n_y : là cỡ mẫu của nhóm y

2.1. Kiểm định trong trường hợp hai mẫu độc lập

(2) Kiểm định trung bình dựa trên giả thuyết là phương sai của hai tổng thể bằng nhau:

Giả thuyết:

. $H_0: \sigma_x^2 = \sigma_y^2$

. $H_1: \sigma_x^2 \neq \sigma_y^2$

Căn cứ vào hai mẫu n_x , n_y được chọn ngẫu nhiên độc lập từ hai tổng thể với phương sai mẫu tương ứng S_x^2 , S_y^2 ta đưa ra quy tắc để quyết định chấp nhận hoặc bác bỏ giả thuyết H_0 với mức ý nghĩa α

2.1. Kiểm định trong trường hợp hai mẫu độc lập

Tính giá trị kiểm định: $\frac{S_x^2}{S_y^2}$

Với giả thuyết: $S_x^2 > S_y^2, S_x^2 \leq S_y^2$

Tra bảng Fisher – Snedecor: $n_x - 1, n_y - 1$

Tìm giá trị: $F_{n_x - 1, n_y - 1, \alpha/2}$

2.1. Kiểm định trong trường hợp hai mẫu độc lập

Nếu:

$$\frac{S_x^2}{S_y^2} \leq F_{n_x-1, n_y-1, \alpha/2} \quad \text{Chấp nhận giả thuyết } H_0$$

$$\frac{S_x^2}{S_y^2} > F_{n_x-1, n_y-1, \alpha/2} \quad \text{Bác bỏ giả thuyết } H_0$$

(SPSS Kiểm định trung bình dựa trên giả thuyết là phương sai của hai tổng thể bằng nhau này)

2.1. Kiểm định trong trường hợp hai mẫu độc lập

Lưu ý: Kiểm định bằng SPSS giá trị kiểm định t , với mức ý nghĩa kiểm định $\alpha = 0,05$ (độ tin cậy 95%). Nếu:

*+ Bác bỏ giả thuyết H_0 khi: **Sig.** < $\alpha = 0,05$*

*+ Chấp nhận giả thuyết H_0 khi: **Sig.** \square $\alpha = 0,05$*

2.1. Kiểm định trong trường hợp hai mẫu độc lập

Thao tác:

- Menu Analyze Compare Means Independent Samples t-Test hộp thoại xuất hiện
- . Test Variable(s): ô đưa biến định lượng
- . Grouping Variable: ô đưa biến định tính (2 giá trị vào)
- . Define Groups: ô đưa biến định tính có thể nhiều hơn 2 giá trị vào (cần xác định nhóm so sánh với nhau)
- . Options: thay đổi độ tin cậy (mặc định là 95%)
 - OK, kết quả kiểm định xuất hiện

2.1. Kiểm định trong trường hợp hai mẫu phụ thuộc

+ Cơ sở lý thuyết:

Mẫu phụ thuộc là mẫu chọn từng cặp tương đồng về các yếu tố tác động. Mỗi đối tượng ở nhóm quan sát này có một đối tượng tương xứng ở nhóm quan sát kia.

Giả sử ta có mẫu n cặp quan sát lấy ngẫu nhiên từ hai tổng thể X và Y : $(y_1, x_1), (y_2, x_2), \dots, (y_n, x_n)$, có phân phối chuẩn. Gọi:

μ_x, μ_y : là trung bình của 2 tổng thể

\bar{D} : là trung bình độ lệch của 2 tổng thể

S_d : là độ lệch chuẩn

2.1. Kiểm định trong trường hợp hai mẫu phụ thuộc

Ta cần kiểm định giả thuyết:

$$H_0: \mu_x - \mu_y = 0 \quad (0 \text{ cho trước})$$

$$H_1: \mu_x - \mu_y \neq 0$$

Dựa vào mẫu với mức ý nghĩa α tra từ bảng phân phối student với $t_{n-1, \alpha/2}$

Quy tắc quyết định:
$$t = \frac{\bar{D}}{\frac{S_d}{\sqrt{n}}}$$

- + Bác bỏ giả thuyết H_0 : $|t| > t_{n-1, \alpha/2}$
- + Chấp nhận giả thuyết H_1 : $|t| \leq t_{n-1, \alpha/2}$

2.1. Kiểm định trong trường hợp hai mẫu phụ thuộc

+ Lưu ý:

Kiểm định bằng SPSS, SPSS tính ra giá trị kiểm định t:

$$t = \frac{\text{Mean}}{\text{Std.Deviation}}$$

Với mức ý nghĩa $\alpha = 0,05$ (độ tin cậy 95%):

- Nếu Sig. < $\alpha = 0,05$: bác bỏ giả thuyết H_0
- Nếu Sig. = $\alpha = 0,05$: : chấp nhận giả thuyết H_0

2.1. Kiểm định trong trường hợp hai mẫu phụ thuộc

Thao tác:

- Menu Analyze Compare Means Paired Samples t-Test hộp thoại xuất hiện
 - . Current Selection: khi chọn biến định lượng hai mẫu phối hợp từng cặp tính trung bình vào sẽ hiện biến lên.
 - . Paired Variables: ô đưa hai biến định lượng hai mẫu phối hợp từng cặp vào.
 - . Options: thay đổi độ tin cậy (mặc định độ tin cậy 95%)
 - OK, kết quả kiểm định xuất hiện

3. Kiểm định phương sai (ANOVA)

+ **Mục đích:** là phân tích ảnh hưởng của một yếu tố nguyên nhân (biến định tính, nhiều giá trị) đến một yếu tố kết quả (biến định lượng)

Thật ra kiểm định phương sai (ANOVA) là kiểm định trung bình (T-Test) lần lượt so sánh các cặp giá trị.

+ **Đối tượng:**

. Biến phụ thuộc: dữ liệu định lượng (tính trung bình)

. Biến độc lập: dữ liệu định tính (so sánh nhiều hơn 2 giá trị)

3. Kiểm định phương sai (ANOVA)

+ Cơ sở lý thuyết:

Giả sử rằng chúng ta muốn so sánh trung bình của k tổng thể có phương sai bằng nhau dựa trên những mẫu ngẫu nhiên độc lập gồm x_1, x_2, \dots, x_k quan sát từ k tổng thể khác nhau có phân phối chuẩn. Nếu trung bình của các tổng thể được kí hiệu là $\mu_1; \mu_2; \dots; \mu_k$ thì mô hình phân tích phương sai một chiều được mô tả dưới dạng kiểm định giả thuyết như sau:

$$H_0: \mu_1 = \mu_2 = \dots = \mu_k$$

Nghĩa là giả thuyết H_0 cho rằng trung bình của k tổng thể khác nhau thì *bằng nhau*. Để kiểm định giả thuyết này cần thực hiện các bước sau:

3. Kiểm định phương sai (ANOVA)

(1) Tổng độ lệch bình phương nội bộ nhóm (SSW):

$$SSW = \sum_{i=1}^k \sum_{j=1}^{n_i} (x_{ij} - x_i)^2$$

(2) Tổng độ lệch bình giữa cá nhóm (SSG):

$$SSG = \sum_{i=1}^k n_i (\bar{x}_i - \bar{x})^2$$

(3) Tổng độ lệch bình phương toàn bộ (SST):

$$SST = \sum_{i=1}^k \sum_{j=1}^{n_i} (x_{ij} - \bar{x})^2$$

3. Kiểm định phương sai (ANOVA)

Bằng các biến đổi toán học ta có: **SST = SSW + SSG**

+ Phương sai nội bộ nhóm: $MSW = \frac{SSW}{n - k}$

+ Phương sai giữa các nhóm: $MSG = \frac{SSG}{k - 1}$

3. Kiểm định phương sai (ANOVA)

Bác bỏ H_0 nếu: $\frac{MSG}{MSW} > F_{k-1, n-k, \alpha}$

Trong đó $F_{k-1, n-k, \alpha}$

là giá trị sao cho $P(F_{k-1, n-k} > F_{k-1, n-k, \alpha}) = \alpha$

$F_{k-1, n-k}$ có phân phối F với bậc tự do của tử số là $(k-1)$ và bậc tự do của mẫu số là $(n-k)$

* Nếu kiểm định bác bỏ H_0 thì phải tiếp tục phân tích sâu ANOVA (thử tực Post Hoc) để xác định sự khác nhau.

3. Kiểm định phương sai (ANOVA)

+ Thao tác:

- Menu Analyze Compare Means One Way ANOVA hộp thoại xuất hiện
 - . Dependent List: ô đưa biến định lượng vào
 - . Factor: ô xác định giá trị so sánh
 - Options hộp thoại xuất hiện
 - . Descriptive: tính các đại lượng thống kê mô tả chi tiết cho từng nhóm phân tích
 - . Homogeneity-of-variance: kiểm định sự bằng nhau của các phương sai nhóm
 - Continue OK. kết quả xuất ANOVA xuất hiện

3. Kiểm định phương sai (ANOVA)

+ Phân tích sâu ANOVA:

(1) Kiểm định trước: là kiểm định các giả định về sự khác nhau của các trung bình nhóm trước khi thực hiện phân tích ANOVA.

Chọn Contrasts ▪ Priori Contrasts.

(2) Kiểm định sau: là kiểm định các giả định về sự khác nhau của các trung bình nhóm sau khi thực hiện phân tích ANOVA.

Chọn Post Hoc ▪ chọn một trong các công cụ kiểm định: LSD; Bonferroni; Tukey; Scheffe; Dunnett

3. Kiểm định phương sai (ANOVA)

(2) Kiểm định sau:

- . LSD: là kiểm định t lần lượt từng cặp trung bình nhóm
- . Bonferroni: là kiểm định giống LSD, nhưng điều chỉnh được mức ý nghĩa kiểm định
- . Tukey: là kiểm định rất hiệu quả khi so sánh trung bình nhiều các cặp giá trị.
- . Dunnett: là kiểm định cho phép chọn so sánh các trị trung bình của các nhóm mẫu còn lại với một nhóm mẫu cụ thể được chọn (SPSS mặc định nhóm cuối cùng)

3. Kiểm định phương sai (ANOVA)

Một số giả định đối với phân tích phương sai một yếu tố:

- các nhóm so sánh phải độc lập và ngẫu nhiên.
- các nhóm so sánh phải có phân phối chuẩn hoặc cỡ mẫu phải đủ lớn, xem như tiệm cận phân phối chuẩn.
- phương sai các nhóm phải đồng nhất

Bài tập: sử dụng tập tin data thuchanh.sav

Kiểm định và bình luận thông số kiểm định:

1. Thu nhập cá nhân [tncn] có tác động đến cách đọc báo SGTT [C15] không?
2. Mức độ thường xuyên đọc báo [C1] có phụ thuộc vào trình độ học vấn [hocvan] không?
3. So sánh trung bình số nhân khẩu [sonk] trong mỗi hộ gia đình ở TP. Hà Nội và TP. Hồ Chí Minh [tp] như thế nào?
4. So sánh trung bình số người xem quảng cáo [24b] giữa nam và nữ [gtingh] như thế nào?
5. So sánh trung bình tuổi [tuoi] của những người làm nghề khác nhau [nghe] như thế nào?
6. So sánh trung bình mức độ đánh giá quan niệm “có nhiều tiền” [c36.1] giữa những người có trình độ học vấn khác nhau [hocvan] như thế nào?

Never say never

Innovation
Differentiation
Reliability **IDR**

UNIVERSITY OF ECONOMICS - HCMC
INSTITUTE OF DEVELOPMENT
ECONOMICS RESEARCH

PHÂN TÍCH DỮ LIỆU VỚI PHẦN MỀM EIEWS

Nguyễn Duy Tâm

Nguyễn Duy Tâm - IDR - Never say never

Never say never

MỤC ĐÍCH – NỘI DUNG

Innovation
Differentiation
Reliability **IDR**

1. Giới thiệu phần mềm Eviews
2. Tạo database phù hợp với dữ liệu –
Nhập liệu
3. Vẽ các loại đồ thị - đặc biệt “đồ thị phân
tán”
4. Những thao tác xử lý cơ bản

Thực hiện với
nhiều cách khác
nhau

Never say never

GIỚI THIỆU PHẦN MỀM EViews



- Là phần mềm thống kê phổ biến với điểm mạnh là xây dựng các mô hình Kinh Tế Lượng



Nguyễn Duy Tâm - IDR - Never say never

3

Never say never

TẠO DATABASE



- Khởi động:
 1. Start/Programs/Eviews
 2. Double click vào biểu tượng

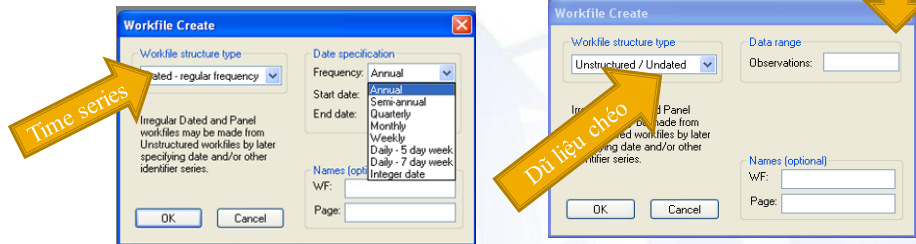


Nguyễn Duy Tâm - IDR - Never say never

4

TẠO DATABASE

- Tạo một file Eviews:
- Phân loại dữ liệu:
- Dữ liệu chuỗi thời gian (annual, quarterly, monthly, daily,...)
 - ➔ Chọn **Dated regular frequency** (chú ý thời gian bắt đầu, thời gian kết thúc)
- Dữ liệu chéo **Unstructured / Undated** (Chú ý số quan sát) n

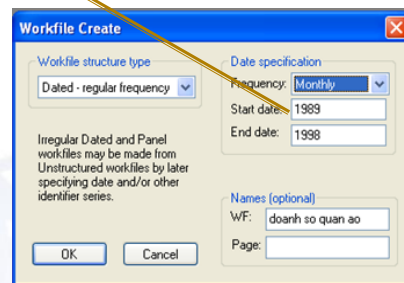


TẠO DATABASE

Dữ liệu chuỗi thời gian
Theo tháng (monthly)
Thời gian bắt đầu: 1-1-1989
Thời gian kết thúc 1-12-1998

date	men	women	jewel	time
1-Jan-1989	11,357.92	16,578.93	10,776.38	1.00
1-Feb-1989	10,605.95	18,236.13	10,821.97	2.00
...
1-Nov-1998	24,289.32	54,617.35	23,753.23	119.00
1-Dec-1998	38,609.66	80,245.97	35,893.86	120.00

Thiết kế trên Eviews



★ Nếu dữ liệu bắt đầu vào tháng 6 -1989 đến tháng 5-1998, Ta làm sao?

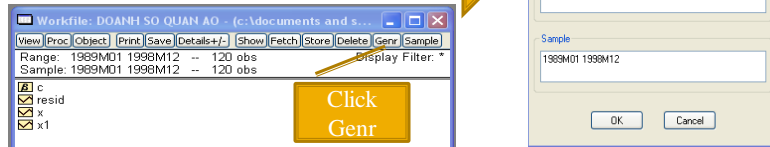
Never say never



TẠO DATABASE – TẠO BIẾN

- Các cách tạo biến trên Eviews

1. Quick/Generate series...



2. Tại vùng nhập lệnh gõ câu lệnh:

a. Genr [space] [tenbiến]=na

VD: genr x=na

b. Series [space] [tenbiến]=na

VD: series x=na

Ý nghĩa câu lệnh: tạo biến X (chưa có dữ liệu)

Lần lược tạo tất cả những biến cần

Nguyễn Duy Tâm - IDR - Never say never

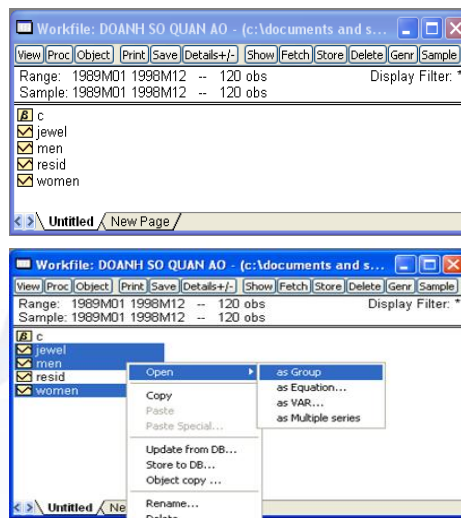
7

Never say never



TẠO DATABASE – Nhập liệu cho biến

- Sau khi đã tạo xong các biến, tiến hành mở nhóm các biến ra một cửa sổ để nhập liệu.
- Mở biến: chọn khối biến → **Right click/open**
- Chú ý:** Biến nào chọn trước sẽ có thứ tự trước



Nguyễn Duy Tâm - IDR - Never say never

8

Never say never



TẠO DATABASE – Nhập liệu

- Các cách nhập liệu:
 1. Dữ liệu trên phiếu: Nhập trực tiếp bằng tay
 2. Dữ liệu trên các phần mềm khác (excel, text,...) có thể copy/past vào cho từng biến.
- **Chú ý:** trước khi nhập liệu, click chọn **Edit+/-**, sau khi nhập xong, chọn lại **Edit+/-**

obs	MEN	WOMEN	JEWEL
1989MD1	NA	NA	NA
1989MD2	NA	NA	NA
1989MD3	NA	NA	NA
1989MD4	NA	NA	NA
1989MD5	NA	NA	NA
1989MD6	NA	NA	NA
1989MD7	NA	NA	NA



Nguyễn Duy Tâm - IDR - Never say never

9

Never say never

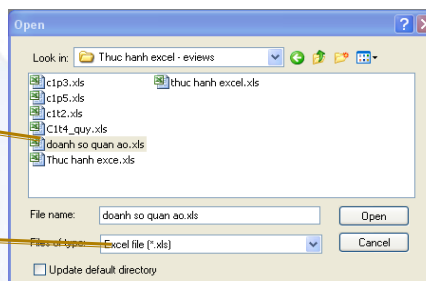
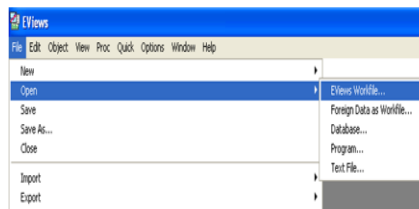


TẠO DATABASE- (mách nước)

- Trong trường hợp bạn đã có một file dữ liệu trên excel hoàn chỉnh, ta có tạo nhanh file evIEWS như sau:
- **Mở trực tiếp evIEWS trên excel**

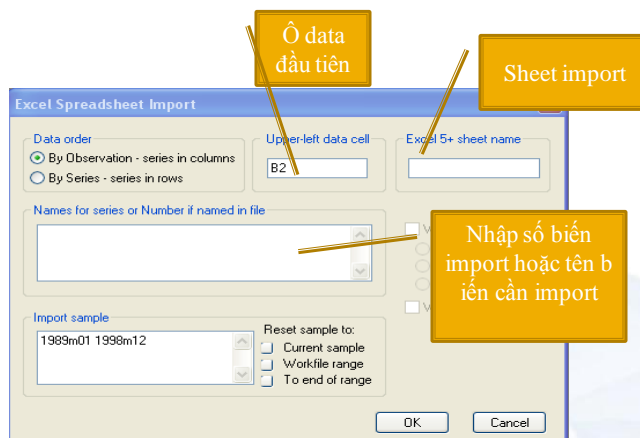
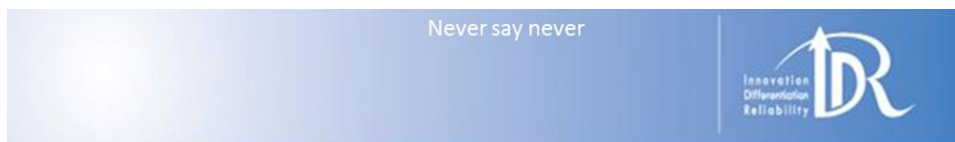
Chọn file excel chứa data

Đổi từ (*.wfl) → (*.xls)

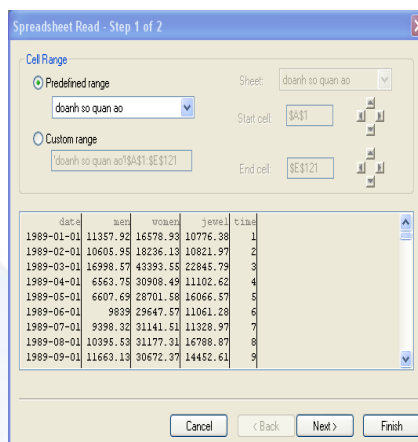


Nguyễn Duy Tâm - IDR - Never say never

10



- Chọn Next/finish
- **Chú ý:** File views mới tạo sẽ ít hiệu chỉnh nếu như file excel đã được chúng ta hiệu chỉnh kỹ (thông thường chỉ có 1 sheet + dữ liệu, không có những vấn đề khác ngoài dữ liệu và tên biến)



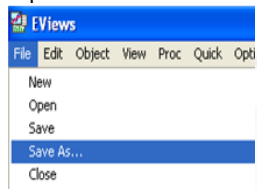
Never say never



TẠO DATABASE- (mách nước)

LƯU FILE

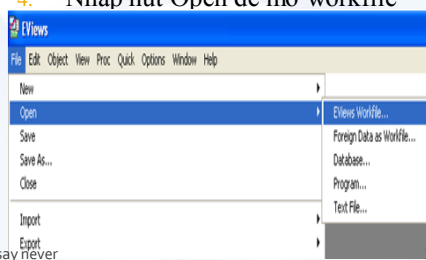
1. File/Save As... hoặc File/Save
2. Chọn Folder chứa workfile tại Save in
3. Đặt tên cho workfile tại File name
4. Chọn kiểu cho workfile tại Save as type (*thông thường ta chọn kiểu *.wfl*)
5. Nhấp nút Save để lưu workfile



Nguyễn Duy Tâm - IDR - Never say never

MỞ FILE

1. Vào thực đơn File/Open/Workfile...
2. Tìm thư mục chứa workfile tại Look in
3. Chọn tên workfile cần mở
4. Nhấp nút Open để mở workfile



Thực hành tạo Database

Never say never



TẠO CÁC FILE SAU

- Tạo các file eviews từ nguồn dữ liệu sẵn có:
- File [doanh so quan ao.xls](#)
- File [Tinh trang cong viec](#)
- File [c1p3](#)
- File [C1p5](#)
- File [c1t2](#)
- File [c1t4 quy](#)

CHÚ Ý

- Tạo theo 2 cách
- C1: tạo biến và copy
- C2: Import

Nguyễn Duy Tâm - IDR - Never say never

14

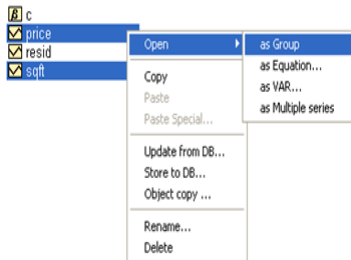
Never say never

NHỮNG THAO TÁC XỬ LÝ CƠ BẢN



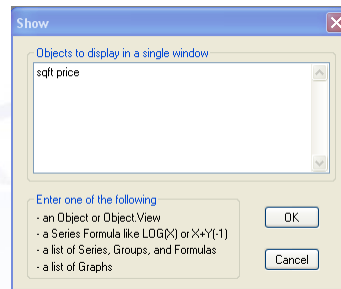
MỞ NHÓM DỮ LIỆU (C1)

- Select biến cần mở (có thể chọn cùng lúc nhiều biến)



MỞ NHÓM DỮ LIỆU (C2)

- Quick/show và gõ tên biến cần mở vào hộp thoại sau



Nguyễn Duy Tâm - IDR - Never say never

15

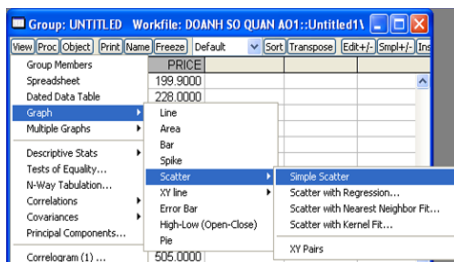
Never say never

NHỮNG THAO TÁC XỬ LÝ CƠ BẢN



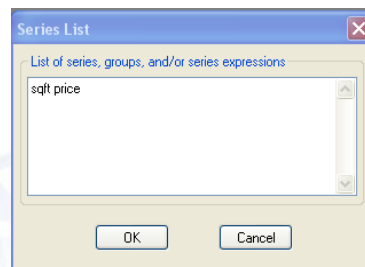
VẼ ĐỒ THỊ PHÂN TÁN (C1)

- Trên cửa **Group** chọn View/Graph/Scatter/simple scatter



VẼ ĐỒ THỊ PHÂN TÁN (C2)

- Quick/graph/scatter



Chú ý, tên biến nào chọn trước làm trục hoành



Nguyễn Duy Tâm - IDR - Never say never

16

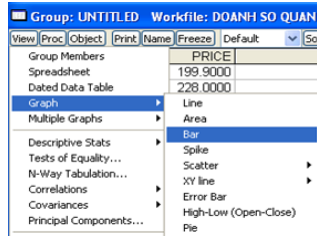
Never say never



NHỮNG THAO TÁC XỬ LÝ CƠ BẢN

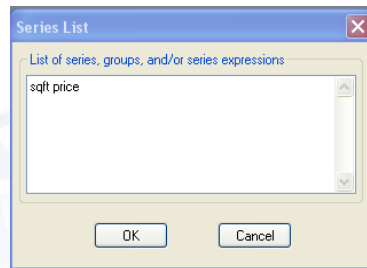
VẼ CÁC LOẠI ĐỒ THỊ KHÁC

- Trên cửa **Group** chọn View/Graph/[tên đồ thị cần vẽ]



VẼ CÁC LOẠI ĐỒ THỊ KHÁC

- Quick/graph/[loại đồ thị cần vẽ]



Thực hành: Vẽ đồ thị line, pie, bar

Nguyễn Duy Tâm - IDR - Never say never

17

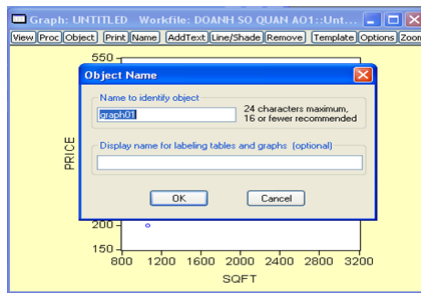
Never say never



NHỮNG THAO TÁC XỬ LÝ CƠ BẢN

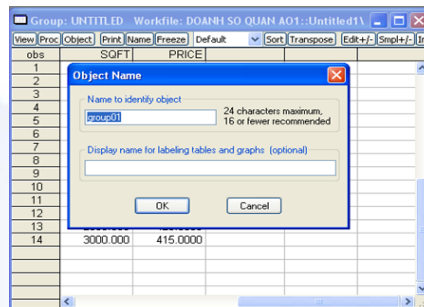
ĐẶT TÊN CHO ĐỒ THỊ

- Từ cửa số Graph, chọn **Name** sau đó đặt tên vào hộp thoại Object Name



ĐẶT TÊN CHO GROUP

- Từ cửa số Group, chọn **Name** sau đó đặt tên vào hộp thoại Object Name



Nguyễn Duy Tâm - IDR - Never say never

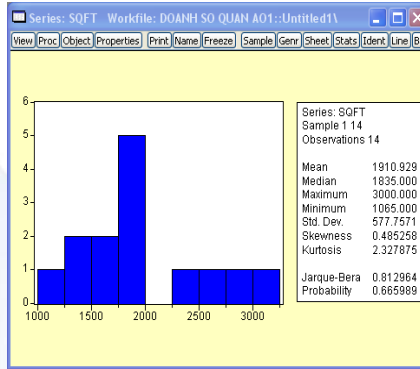
18

ĐỒ THỊ HISTOGRAM & STAT

- Quick/series statistics/histogram and stats



ĐỒ THỊ HISTOGRAM & STAT



MỘT SỐ LỆNH CƠ BẢN

1. Trung bình: @mean(tên biến)
2. Trung vị: @media(tên biến)
3. Phương sai: @var(tên biến)
4. Độ lệch chuẩn: @stdev(tên biến)
5. Bách phân vị: @quantile(tên biến)
6. Hệ số tương quan: @cor(tên biến)
7. Tương quan: @cov(tên biến)
8. Số quan sát: @obs(tên biến)

ỨNG DỤNG CHO BIẾN SQFT, PRICE

1. Genr xtb=@mean(sqft)
2. Genr Xme=@median(sqft)
3. Genr Xps=@var(sqft)
4. Genr Xstd=@stdev(sqft)
5. Genr X50=@quantile(sqft)
6. Genr Xcor=@cor(sqft)
7. Genr Xcov=@cov(sqft)
8. Genr Xobs=@obs(sqft)

Chú ý: Tất cả các lệnh đều trực tiếp hiện ở vùng nhập lệnh

Never say never

NHỮNG THAO TÁC XỬ LÝ CƠ BẢN



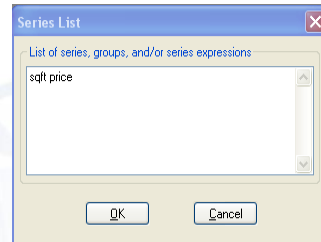
TÍNH NHANH CÁC GIÁ TRỊ THỐNG KÊ MÔ TẢ

- Từ cửa sổ group:
- View/descriptive stats/common samle

	SQFT	PRICE
Mean	1910.929	317.4529
Median	1835.000	291.5000
Maximum	3000.000	605.0000
Minimum	1065.000	199.9000
Std. Dev.	577.7571	88.49816
Skewness	0.485258	0.653457
Kurtosis	2.327875	2.470167
Jarque-Bera	0.812964	1.160102
Probability	0.665989	0.559870
Sum	26753.00	4444.900
Sum Sq. Dev.	4339443.	101815.0
Observations	14	14

TÍNH NHANH CÁC GIÁ TRỊ THỐNG KÊ MÔ TẢ

- Quick/group statistics/descriptive statistics/common sample.
- Nhập tên biến cần tính



Nguyễn Duy Tâm - IDR - Never say never

21

Never say never

NHỮNG THAO TÁC XỬ LÝ CƠ BẢN



LẬP MA TRẬN HỆ SỐ TƯƠNG QUAN

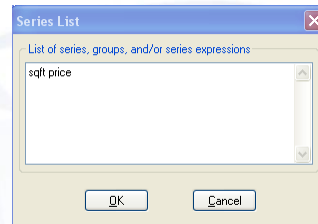
1. B1: Mở những biến cần tính hệ số tương quan
2. B2: Từ cửa sổ group: View/correlation/common sample

	SQFT	PRICE
SQFT	1.000000	0.905827
PRICE	0.905827	1.000000

Tương tự cho việc lập ma trận tương quan (Hiệp phương sai)

LẬP MA TRẬN HỆ SỐ TƯƠNG QUAN

1. Quick/group statistics/correlations
2. Nhập tên những biến cần tính hệ số tương quan



Nguyễn Duy Tâm - IDR - Never say never

22

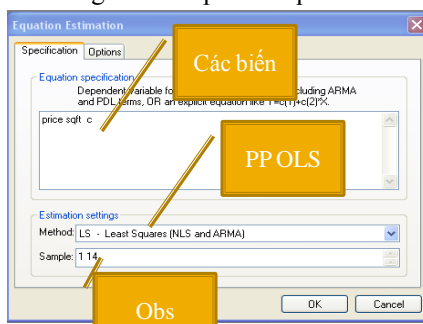
Never say never



NHỮNG THAO TÁC XỬ LÝ CƠ BẢN

LẬP MÔ HÌNH HỒI QUY ĐƠN BIẾN

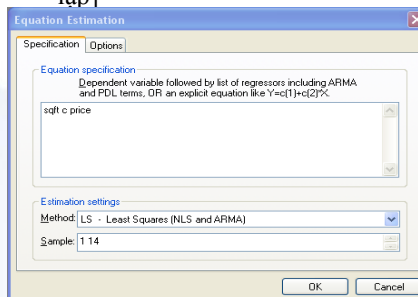
1. Chọn lần lượt biến phụ thuộc, biến độc lập
2. Right click/open/as equation...



Nguyễn Duy Tâm - IDR - Never say never

LẬP MÔ HÌNH HỒI QUY ĐƠN BIẾN

1. Quick/estimate equation
2. Nhập lần lượt [biến phụ thuộc] [space] [C] [space] [biến độc lập]



23

Never say never



NHỮNG THAO TÁC XỬ LÝ CƠ BẢN

1. **Dependent Variable:** Biến phụ thuộc
2. **Method:** Phương pháp ước lượng
3. **Sample:** Số quan sát của mẫu
4. **Included observation:** Số quan sát ước lượng
5. **Variable:** Cột biến độc lập và hằng số
6. **Coefficient:** Cột hệ số hồi quy
7. **Std.Error:** Cột sai số chuẩn
8. **T_Statistics:** Cột thống kê t
9. **Prob:** Cột giá trị Prob
10. **R-square:** Giá trị R bình phương
11. **F-statistic:** Cột thống kê F
12. **Prob(f-statistic):** Cột prob của F

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	52.35091	37.28549	1.404056	0.1857
SQFT	0.138750	0.018733	7.406788	0.0000

R-squared	0.820522	Mean dependent var	317.4929
Adjusted R-squared	0.805565	S.D. dependent var	88.49816
S.E. of regression	39.02304	Akaike info criterion	10.29774
Sum squared resid	18273.57	Schwarz criterion	10.38904
Log likelihood	-70.08421	F-statistic	54.86051
Durbin-Watson stat	1.975057	Prob(F-statistic)	0.000008

Nguyễn Duy Tâm - IDR - Never say never

24

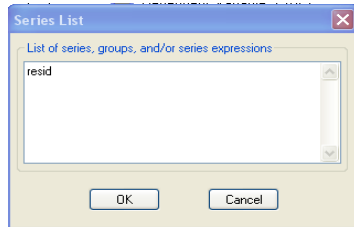
Never say never

NHỮNG THAO TÁC XỬ LÝ CƠ BẢN



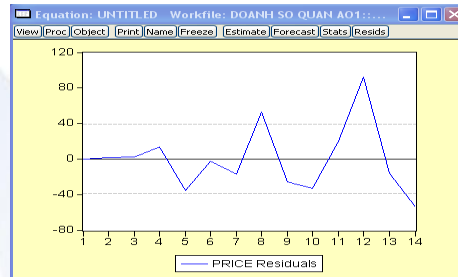
ĐỒ THỊ PHẦN DƯ (RESID)

1. Sau khi đánh giá mô hình, ta có thể vẽ đồ thị phần dư:
2. Quick/graph/line graph



ĐỒ THỊ PHẦN DƯ (RESID)

1. Từ cửa sổ Equation
2. View/Actual, fitted/residual/residual graph.



Nguyễn Duy Tâm - IDR - Never say never

25

Never say never

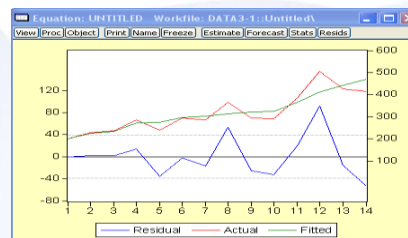
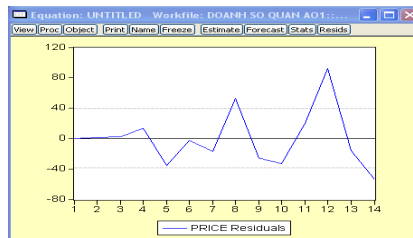
NHỮNG THAO TÁC XỬ LÝ CƠ BẢN



Từ cửa sổ workfile

1. View/Actual, Fitted, residual/residual graph.
2. View/Actual, Fitted, residual/Actual, Fitted, residual graph

Tạo mô hình theo cách "Pro" theo lệnh
Equation [tên phương trình].ls
[biến phụ thuộc] c [biến độc lập]



Nguyễn Duy Tâm - IDR - Never say never

26

Never say never

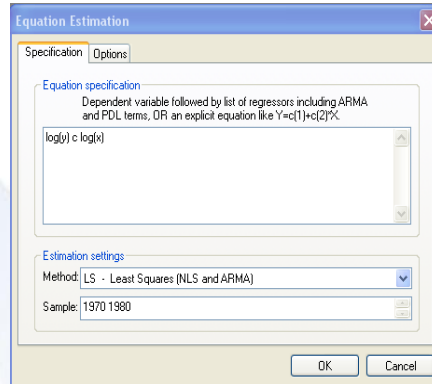
NHỮNG THAO TÁC XỬ LÝ CƠ BẢN



TẠO MÔ HÌNH LOG - LOG

- Dùng bài tập *thidu3_3.wf1* dựng mô hình sau:
 $LOG(Y) = \beta_0 + \beta_1 * LOG(X)$.
- Cách 1: Tại hộp thoại Equation Estimation.**
 $Log(y) \text{ c } Log(x)$
- Cách 2: dùng lệnh**
Equation log_log.ls log(y) c log(x)
- Cách 3: Tạo các biến**
 $log(y), Log(x)$

TẠO MÔ HÌNH LOG - LIN



Nguyễn Duy Tâm - IDR - Never say never

27

Never say never

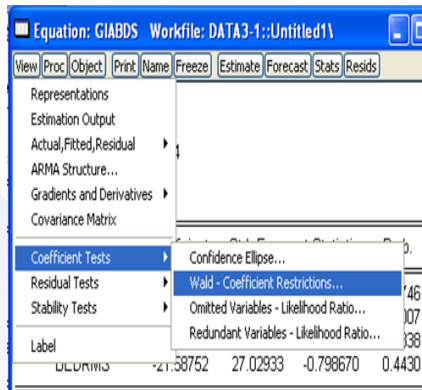
NHỮNG THAO TÁC XỬ LÝ CƠ BẢN



KIỂM ĐỊNH WALD

- Kiểm định đồng thời sự bằng 0 của các hệ số hồi quy.
 $\beta_i = \beta_j = 0$
- Từ cửa sổ Equation, thực hiện theo đường link bên

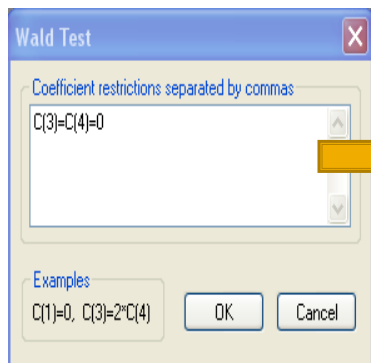
KIỂM ĐỊNH WALD



Nguyễn Duy Tâm - IDR - Never say never

28

Never say never

NHỮNG THAO TÁC XỬ LÝ CƠ BẢN**KIỂM ĐỊNH WALD****KIỂM ĐỊNH WALD**

Equation: UNTITLED Workfile: DATA4-1.:Untitled

View Proc Object Print Name Freeze Estimate Forecast Stats Resids

Wald Test:
Equation: Untitled

Test Statistic	Value	df	Probability
F-statistic	0.471106	(2, 10)	0.6375
Chi-square	0.942211	2	0.6243

Null Hypothesis Summary:

Normalized Restriction (= 0)	Value	Std. Err.
C(3)	-12.19276	43.25000
C(4)	-21.58752	27.02933

Restrictions are linear in coefficients.



Nguyễn Duy Tâm - IDR - Never say never

29

Never say never

NHỮNG THAO TÁC XỬ LÝ CƠ BẢN**KIỂM ĐỊNH THỪA BIẾN**

Nguyễn Duy Tâm - IDR - Never say never

30

CHƯƠNG 5

PHÂN TÍCH DỮ LIỆU VỚI MICROSOFT EXCEL

TỔNG QUAN

- ❖ Phân tích thống kê
- ❖ Kiểm định giả thiết thống kê
- ❖ Công cụ Add-in Analysis
- ❖ Quy hoạch tuyến tính
- ❖ Kiểm soát chất lượng

PHÂN TÍCH THỐNG KÊ

❖ Đồ thị Box Plot

Đồ thị là một phương pháp rất phổ biến trong thống kê mô tả dùng để rút ra các nhận xét trực quan ban đầu về phân bố dữ liệu.

Ngày	1	2	3	4	5	6	7	8	9	10	11
Giá	10	11	10	11	11	12	12	13	14	16	15

Ngày	12	13	14	15	16	17	18	19	20	21	22
Giá	11	18	19	20	15	14	14	22	25	27	23

Ngày	23	24	25	26	27	28	29	30	31	32	33
Giá	22	26	27	29	28	31	32	30	32	34	33

Ngày	34	35	36	37	38	39	40	41	42	43	44
Giá	38	41	40	42	53	52	47	37	23	11	32

PHÂN TÍCH THỐNG KÊ

Đồ thị Box Plot (tt)

Người phân tích thực hiện các bước như sau:

- Nhập dữ liệu giá vào ô B3:AS3 của bảng tính.
- Nhập công thức vào các ô B9:B13 như trong bảng sau

Ô	Ý nghĩa
B9	Tìm điểm tứ phân vị Q1
B10	Tìm giá trị nhỏ nhất
B11	Tìm điểm trung vị
B12	Tìm giá trị lớn nhất
B13	Tìm điểm tứ phân vị Q3

PHÂN TÍCH THỐNG KÊ

❖ Đồ thị Box Plot (tt)

- Chọn menu *Insert/ Chart....* Trong hộp thoại *Chart Type* chọn *Line*.
- Trong hộp thoại *Chart Source Data* nhập A8:B13 vào mục *Data range* (hay bấm nút để chọn trực tiếp ô chứa dữ liệu).
- Trong hộp thoại *Chart Options* nhập dòng chữ “Đồ thị box plot” vào mục *Chart Title*.
- Trong hộp thoại *Chart Location* chọn mục *As object in* nếu muốn chèn đồ thị vào cùng bảng tính dữ liệu hay chọn mục *As new sheet* nếu muốn tạo bảng tính mới cho đồ thị. Nhấn nút *Finish* để Excel hoàn thành đồ thị.
- Trong đồ thị do Excel tạo ra, nhấn chuột phải lên chuỗi dữ liệu (*Data series*) để hiện thị popup menu, chọn *Format Data Series*. Trong hộp thoại chọn thẻ *Options*, đánh dấu mục *High-low lines* và *Up-down bar*.

PHÂN TÍCH THỐNG KÊ

❖ Công cụ Add-in Data Analysis

cụ Data Analysis của Excel cho phép người sử dụng thực hiện các phân tích thống kê dữ liệu thường gặp. Việc sử dụng Data Analysis rất đơn giản, bằng cách chọn phân tích mà cần thực hiện trong hộp thoại Data Analysis. Dưới đây là liệt kê các phân tích thống kê có thể thực hiện:

- *ANOVA*: phân tích phương sai.
- *Correlation*: phân tích tương quan.
- *Covariance*: phân tích hiệp biến.
- *Descriptive Statistic*: phân tích thống kê mô tả.

PHÂN TÍCH THỐNG KÊ

- *Exponential Smoothing*: dự báo bằng phương pháp làm trơn hàm mũ.
- *F-test Two samples for Variances*: kiểm định F, giả thiết về hai phương sai.
- *Fourier Analysis*: phân tích Fourier.
- *Histogram*: lập biểu đồ histogram.
- *Moving Average*: dự báo bằng phương pháp trung bình dịch chuyển.
- *Random Number Generation*: tạo các số ngẫu nhiên.
- *Rank and Percentile*: thực hiện xếp hạng dữ liệu.
- *Regression*: phân tích hồi quy.

PHÂN TÍCH THỐNG KÊ

- *Sampling*: lấy mẫu từ một tập hợp.
- *t-Test Paired Two Sample For Means*: thực hiện kiểm định giả thiết về trị trung bình của 2 phân bố, lấy mẫu theo cặp.
- *t-Test Two Samples Assuming Equal Variances*: kiểm định t giả thiết về trị trung bình của 2 phân bố có phương sai (chưa biết) bằng nhau.
- *t-Test Two Samples Assuming Unequal Variances*: kiểm định giả thiết về trị trung bình, của 2 phân bố có phương sai (chưa biết) khác nhau.
- *z-Test Two Sample For Means*: kiểm định giả thiết về trị trung bình 2 phân bố với phương sai cho trước.

PHÂN TÍCH THỐNG KÊ

- Ví dụ: Hai phương pháp sản xuất vỏ xe được thử nghiệm để xác định tính năng. 10 vỏ xe theo phương pháp 1 và 8 vỏ xe theo phương pháp 2. Vỏ xe chế tạo theo phương pháp 1 được thực nghiệm tại địa điểm A và theo phương pháp 2 được thực nghiệm tại địa điểm B. Số liệu quá khứ cho thấy rằng tuổi thọ vỏ xe có phân bố normal và độ lệch chuẩn thay đổi tùy theo địa điểm thực nghiệm. Biết độ lệch chuẩn cho địa điểm A là 4000 km và địa điểm B là 6000 km. Nhà sản xuất muốn kiểm định giả thiết là 2 phương pháp đều cho vỏ xe có tuổi thọ như nhau với mức ý nghĩa 5%. Số liệu tuổi thọ vỏ xe trong bảng sau:

PHÂN TÍCH THỐNG KÊ

Địa điể m	Tuổi thọ, x 100km									
A	61,1	58,2	62,3	64,0	59,7	66,2	57,8	61,4	62,2	63,6
B	62,2	56,6	66,4	56,2	57,4	58,4	57,6	65,4		

PHÂN TÍCH THỐNG KÊ

- Nhập dữ liệu như bảng vào các ô A1:K3 của bảng tính.
- Chọn menu *Tools/ Data Analysis*. Trong hộp thoại *Data Analysis* chọn mục *z-Test: Two Sample for Means*.
- Trong hộp thoại *z-Test*, nhập B2:K2 vào mục *Variable 1 Range*, và B3:I3 vào mục *Variable 2 Range*. Nhập giá trị 0 vào mục *Hypothesized Mean Difference* (kiểm định giả thiết). Nhập giá trị phương sai 1600 cho địa điểm A vào mục *Variable 1 Variance* (độ lệch chuẩn được chia 100 để đơn giản dữ liệu). Nhập giá trị phương sai 3600 cho địa điểm B vào mục *Variable 2 Variance*.
- Nhập giá trị mức ý nghĩa 0.05 vào mục *Alpha*, nhấn *OK* để hoàn thành.

TỐI ƯU

❖ Solver Options

Excel có thể giải các bài toán quy hoạch tối ưu tuyến tính, quy hoạch nguyên và quy hoạch phi tuyến bằng cách sử dụng công cụ *Solver*. Công cụ *Solver* cung cấp các tùy chọn như sau:

- *Max time*: giới hạn thời gian mà *Solver* chạy để giải các bài toán. Giá trị 100s là đủ cho hầu hết các bài toán.
- *Iterations*: giới hạn số vòng lặp của *Solver*. Giá trị 100 là đủ cho hầu hết các bài toán.
- *Precision*: thiết lập sai số cho ràng buộc. Giá trị trong khoảng 0...1.

TỐI ƯU

- *Tolerance*: thiết lập sai số cho ô hàm mục tiêu trong bài toán quy hoạch nguyên.
- *Convergence*: thiết lập sai số cho ô mục tiêu trong bài toán quy hoạch phi tuyến. Giá trị từ 0..1
- *Assume Linear Model*: chọn khi giải bài toán quy hoạch tuyến tính.
- *Assume Non-Negative*: chọn khi các biến có giá trị không âm
- *Use Automatic Scaling*: chọn khi khác biệt giữa giá trị của kết quả và dữ liệu là rất lớn.
- *Show Iteration Results*: Solver dừng sau mỗi vòng lặp và hiển thị kết quả của vòng lặp.

TỐI ƯU

- *Estimates*: chọn phương pháp mà *Solver* dùng ước lượng các giá trị ban đầu của lời giải. Chọn *Quadratic* khi giải bài toán phi tuyến.
- *Derivatives*: chọn phương pháp tính đạo hàm mà *Solver* dùng để giải toán. Chọn *Forward* trong hầu hết các bài toán. Chọn *Central* khi các giá trị hàm ràng buộc thay đổi rất lớn.
- *Search*: chọn thuật toán mà *Solver* dùng để tìm lời giải. *Newton* đòi hỏi nhiều bộ nhớ như chạy ít vòng lặp. Phương pháp *Conjugate* cần ít bộ nhớ như chạy nhiều vòng lặp.

TỐI ƯU

➤ Ví dụ:

Một nhà máy có thể sản xuất 3 sản phẩm khác nhau A,B,C. Mỗi sản phẩm yêu cầu phải được gia công trên các máy 1,2,3 theo bảng dưới. Mỗi máy chạy 8 giờ. Giá bán của A,B,C là 5000,4000,5000 đồng. Chi phí vật tư cho A,B,C là 4000,1000,3000 đồng. Chi phí chạy máy 1,2,3 là 1000,2000,1000 đồng/ giờ. Tìm số lượng sản phẩm A,B, C sản xuất mỗi ngày sao cho lợi nhuận thu được là tối đa

TỐI ƯU

Sản phẩm	Máy , phút/đơn vị		
	1	2	3
A	10	8	5
B	7	9	10
C	5	7	6

Nếu ký hiệu A, B, C là số lượng sản phẩm A, B, C sản xuất trong một ngày; X, Y, Z là số giờ chạy máy 1, 2, 3 trong một ngày. Ta có mô hình toán sau:

Hàm mục tiêu:

$$\text{MAX } 1000 A + 3000 B + 2000 C - 1000 X - 2000 Y - 1000 Z$$

Ràng buộc:

$$10 A + 7 B + 5 C - 60 X = 0$$

$$8 A + 9 B + 7 C - 60 Y = 0$$

$$5 A + 10 B + 6 C - 60 Z = 0$$

$$X < 8$$

$$Y < 8$$

$$Z < 8$$

	A	B	C	D	E	F	G
1	Bài toán kế hoạch sản xuất						
2	Dữ liệu						
3	Biến số	Hệ số					
4		A	B	C	X	Y	Z
5	Hàm mục tiêu	1000	3000	2000	-1000	-2000	-1000
6	Ràng buộc 1	10	7	5	-60	0	0
7	Ràng buộc 2	8	9	7	0	-60	0
8	Ràng buộc 3	5	10	6	0	0	-60
9	Ràng buộc 4	0	0	0	1	0	0
10	Ràng buộc 5	0	0	0	0	1	0
11	Ràng buộc 6	0	0	0	0	0	1

Sheet1 / Sheet2 / Sheet3

	A	B	C	D
14	Mô hình toán			
15	Tên	LHS	Yêu cầu	RHS
16	Hàm mục tiêu	0	MAX	
17	Ràng buộc 1	0 =		0
18	Ràng buộc 2	0 =		0
19	Ràng buộc 3	0 =		0
20	Ràng buộc 4	0 ≤		8
21	Ràng buộc 5	0 ≤		8
22	Ràng buộc 6	0 ≤		8

Sheet1 / Sheet2 / Sheet3

	A	B	C	D	E	F	G
24	Kết quả						
25							
26	Biến số	A	B	C	X	Y	Z
27	Giá trị						

Ô	Công thức
B16	=\$B\$27*B5+\$C\$27*C5+\$D\$27*D5+\$E\$27*E5+\$F\$27*F5+\$G\$27*G5
B17	=\$B\$27*B6+\$C\$27*C6+\$D\$27*D6+\$E\$27*E6+\$F\$27*F6+\$G\$27*G6
B18	=\$B\$27*B7+\$C\$27*C7+\$D\$27*D7+\$E\$27*E7+\$F\$27*F7+\$G\$27*G7
B19	=\$B\$27*B8+\$C\$27*C8+\$D\$27*D8+\$E\$27*E8+\$F\$27*F8+\$G\$27*G8
B20	=\$B\$27*B8+\$C\$27*C8+\$D\$27*D8+\$E\$27*E8+\$F\$27*F8+\$G\$27*G9
B21	=\$B\$27*B10+\$C\$27*C10+\$D\$27*D10+\$E\$27*E10+\$F\$27*F10+\$G\$27*G10
B22	=\$B\$27*B11+\$C\$27*C11+\$D\$27*D11+\$E\$27*E11+\$F\$27*F11+\$G\$27*G11

Solver Parameters



Set Target Cell:

Solve

Equal To:

Max

Min

Value of:

Close

By Changing Cells:

Guess

Subject to the Constraints:

Add

Delete

Options

Reset All

Help

Add Constraint



Cell Reference:

Constraint:

OK

Cancel

Add

Help

CÁC HÀM ĐIỀU KIỆN

- IF(Logic,"True Value","False Value")
- IF(AND(Condition 1,Condition 2),"True Value","False Value")
- IF(OR(Condition 1,Condition 2),"True Value","False Value")
- IF(Logic1,"True Value1", IF(Logic2,"True Value2", IF(Logic3,"True Value3",..., "False Value")...))
- COUNTIF(Range,Criteria)

TÌM ĐIỂM HOÀ VỐN

- Tổng chi phí (TC) $TC = cf + vcv$
- Tổng doanh thu (TR) $TR = v.p$
- Lợi nhuận (Z) $Z = TR - TC = v.p - (cf + vcv)$
- trong đó: cf - chi phí cố định
- v - số lượng SP đã được bán
- cv - chi phí biến đổi cho từng đơn vị SP
- p - đơn giá bán của SP
- Điểm hòa vốn được xác định bởi công thức:
• **$TR = TC \rightarrow vp = (cf + vcv)$**

• \rightarrow

$$v \frac{cf}{p - v}$$

TÌM ĐIỂM HOÀ VỐN

- Ví dụ:

1	Chi phí biến đổi 1 sp (C_v)	2700
2	Chi phí cố định (C_f)	650000
3	Giá bán 1 sp (p)	3400
4	Sản lượng sản xuất trong tháng (v)	
5	Doanh thu trong tháng (TR)	
6	Tổng chi phí trong tháng (TC)	
7	Lợi nhuận (Z)	



Hàm mục tiêu *GoalSeek*

KINH TẾ KỸ THUẬT

- Đánh giá hiệu quả kinh tế của dự án là một phần việc quan trọng của kỹ sư kỹ thuật hệ thống công nghiệp. Việc tính toán các giá trị tương đương của dòng tiền tệ, hệ số chiết khấu hay lãi suất nội suy thường đơn điệu và dễ sai sót. Để tránh sai sót và tiết kiệm thời gian áp dụng công cụ bảng tính là điều bắt buộc. Excel cung cấp cho người sử dụng một số công thức tài chính rất hữu dụng trong đánh giá hiệu quả kinh tế của dự án kỹ thuật.

KINH TẾ KỸ THUẬT

❖ Xác định NPV và IRR để đánh giá hiệu quả đầu tư:

- Để xác định hiệu quả đầu tư, người ta thường sử dụng hai chỉ tiêu NPV và IRR. Nếu NPV dương và IRR lớn hơn lãi suất ngân hàng tại thời điểm lập dự án thì dự án có hiệu quả về mặt tài chính
- Ví dụ: Dự án đầu tư X, sau khi phân tích tính toán, các thông số được xây dựng sau đây:

Số vốn đầu tư ban đầu = 15000, vòng đời dự án là 6 năm

Chi phí sử dụng vốn là 12%/năm (lãi suất tiền gửi Ngân hàng)

Khoảng thu hồi qua các năm như sau:

	Năm 0	Năm 1	Năm 2	Năm 3	Năm 4	Năm 5
Vốn đầu tư	15000					
Dòng tiền		2000	4000	5500	6500	4500

KINH TẾ KỸ THUẬT

- NPV được tính theo công thức = Vốn đầu tư (Năm 0) + Hàm NPV
- Trong hàm NPV thì các giá trị Value được tính bắt đầu từ năm 1 là năm bắt đầu đi vào hoạt động
- Khi tính IRR, thì các giá trị Value được tính luôn cả năm 0 là năm đầu tư.