



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**

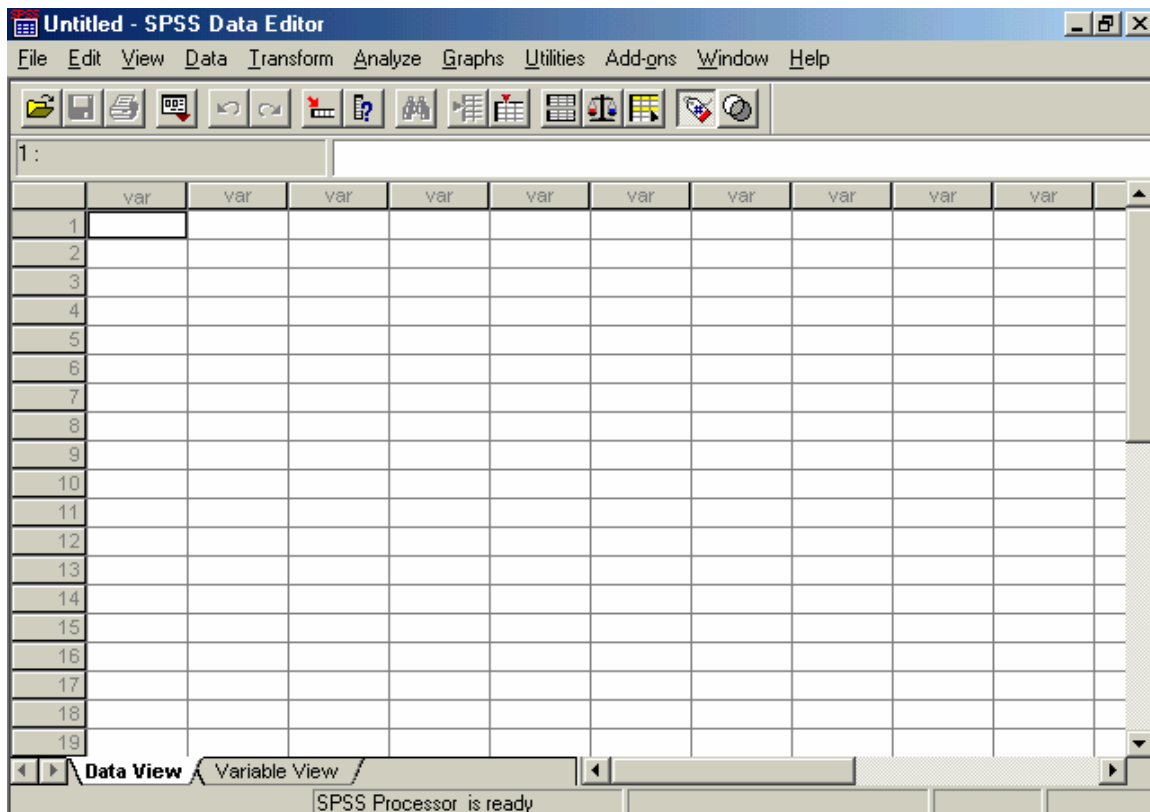
* SPSS là một nhãn hiệu đã đăng ký bản quyền của SPSS Inc.

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

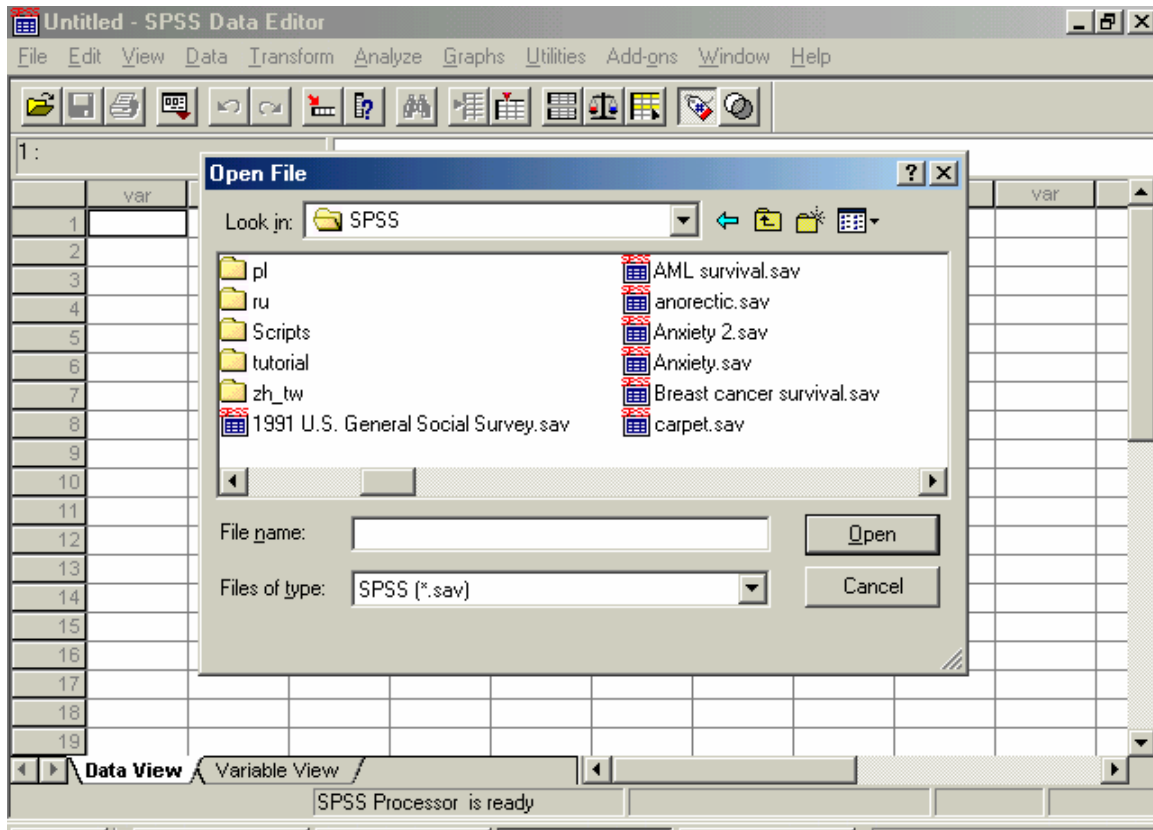
- File:** Chứa nội dung về: Tạo mới dữ liệu, Mở dữ liệu từ các tập tin của SPSS và của các phần mềm khác, Lưu trữ dữ liệu, In ấn và Thoát khỏi SPSS.
- Edit:** Chứa nội dung về: Thực hiện lại hoặc hủy bỏ một thao tác vừa mới thực hiện, Sao chép và cắt dán dữ liệu, Chỉnh sửa một số nội dung tùy chọn chung (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**

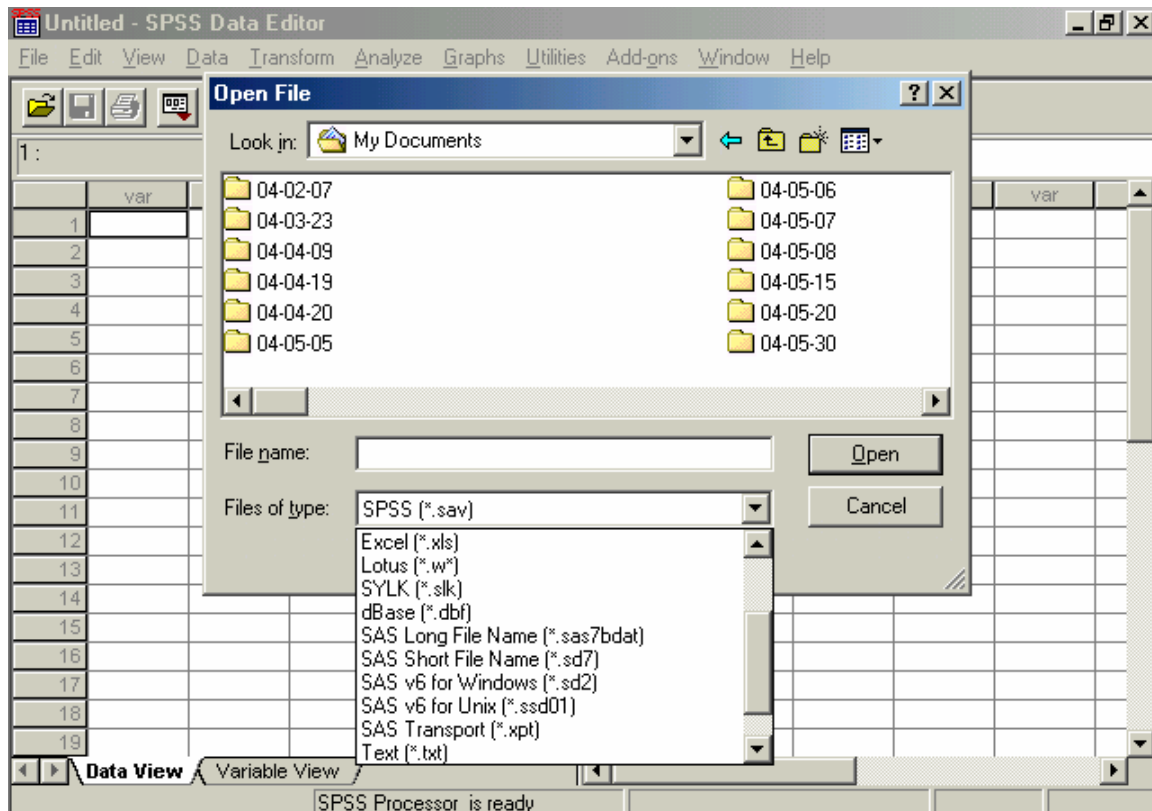
HÌNH 2



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:

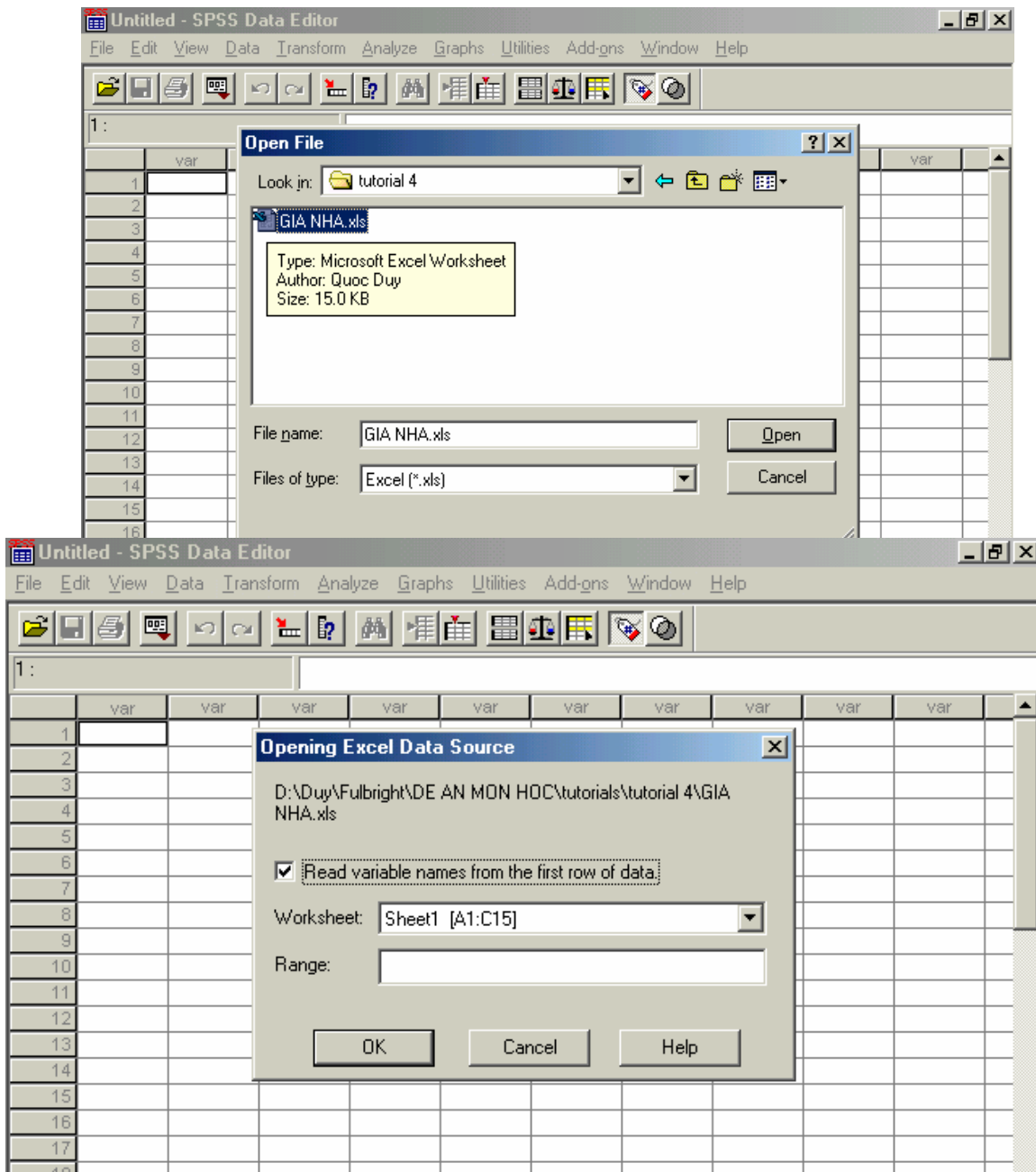
Có 1 tập tin EXCEL chứa dữ liệu về giá nhà và diện tích như đã học. Nội dung của tập tin này bắt đầu từ A1 đến C15. Bây giờ chuyển nội dung của tập tin này sang SPSS.

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.

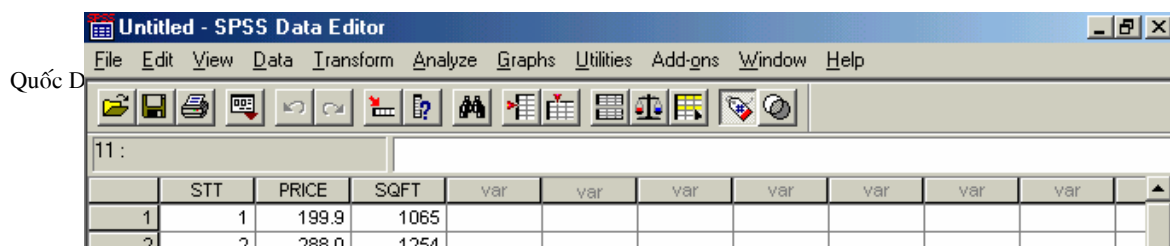
HÌNH 5



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.

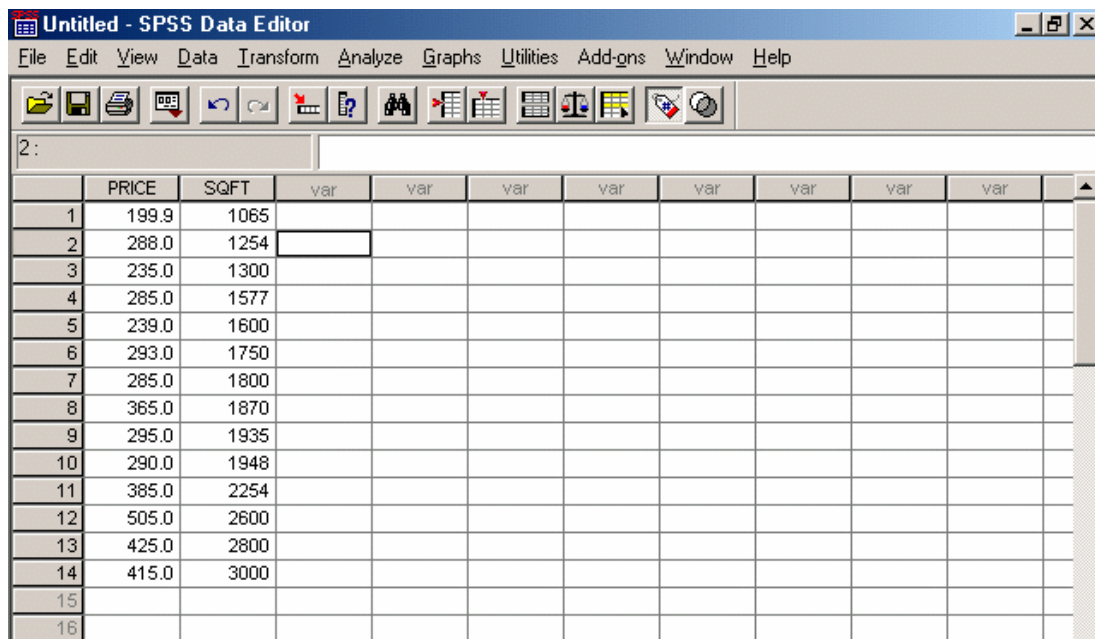
- Nếu chúng ta lấy hết toàn bộ nội dung của file Excel thì nhấn **OK**.

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 16 rows and 11 columns. The first two columns are labeled PRICE and SQFT. The remaining columns are labeled 'var'. The data for the first 14 rows is as follows:

	PRICE	SQFT	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 '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

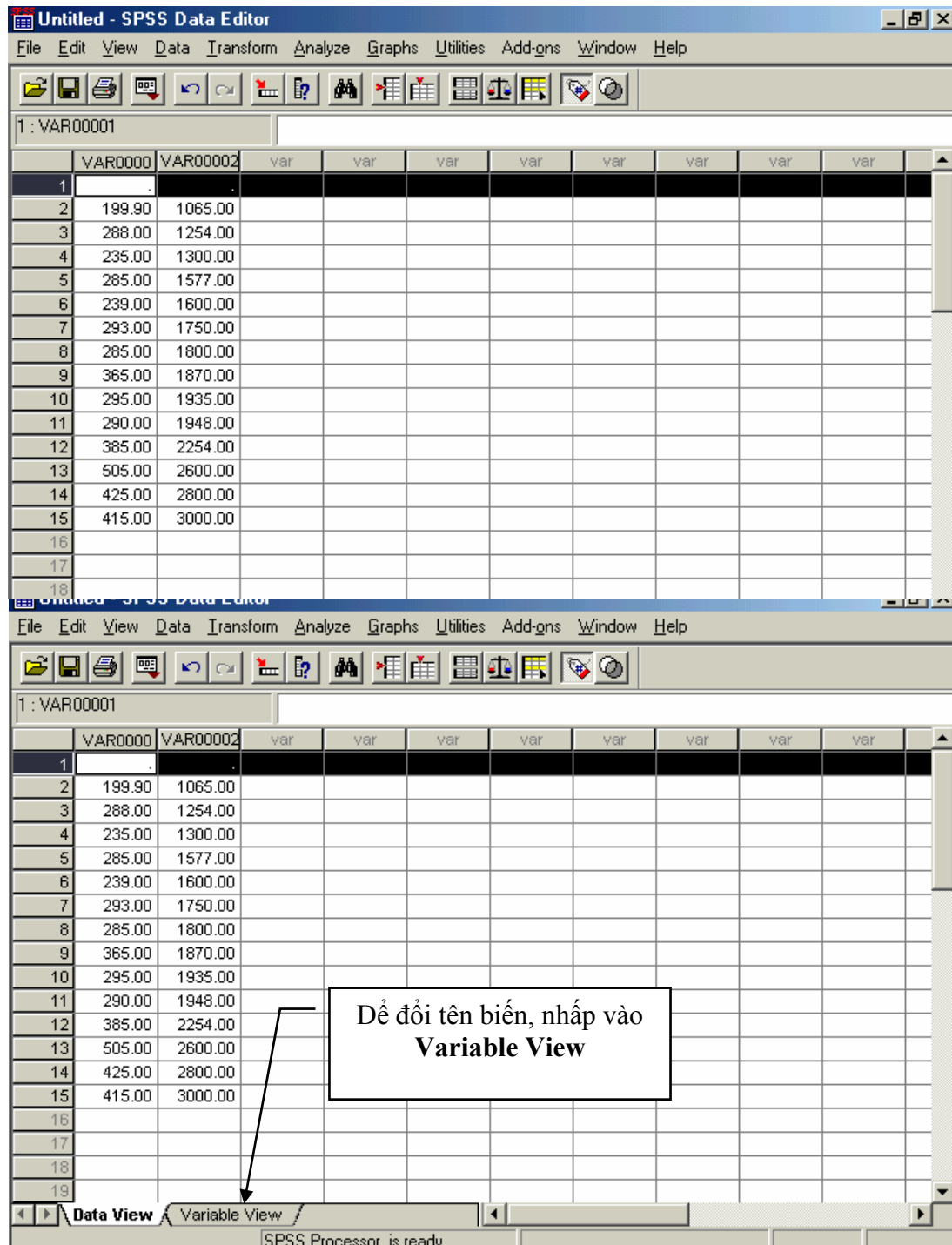
The screenshot shows the SPSS Data Editor window titled 'Untitled - SPSS Data Editor'. The data is imported into a grid with the following structure:

	VAR00001	VAR00002	var	var	var	var	var	var	var	var
1	.	.								
2	199.90	1065.00								
3	228.00	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										

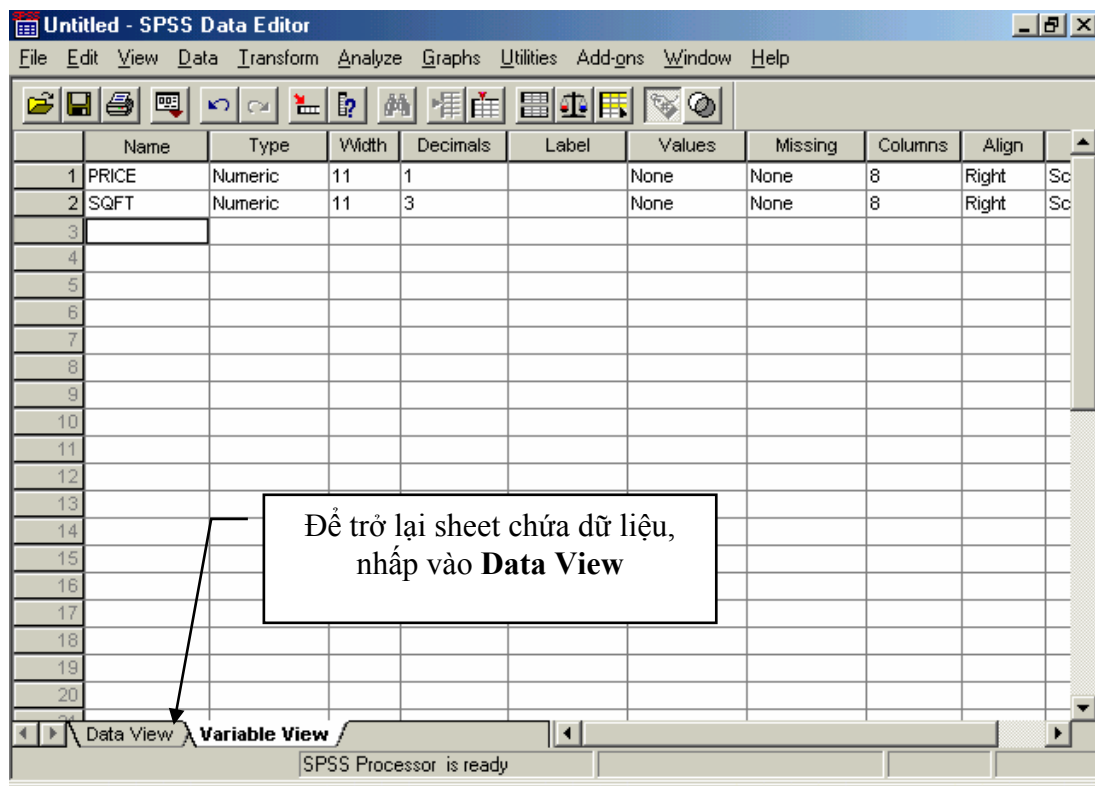
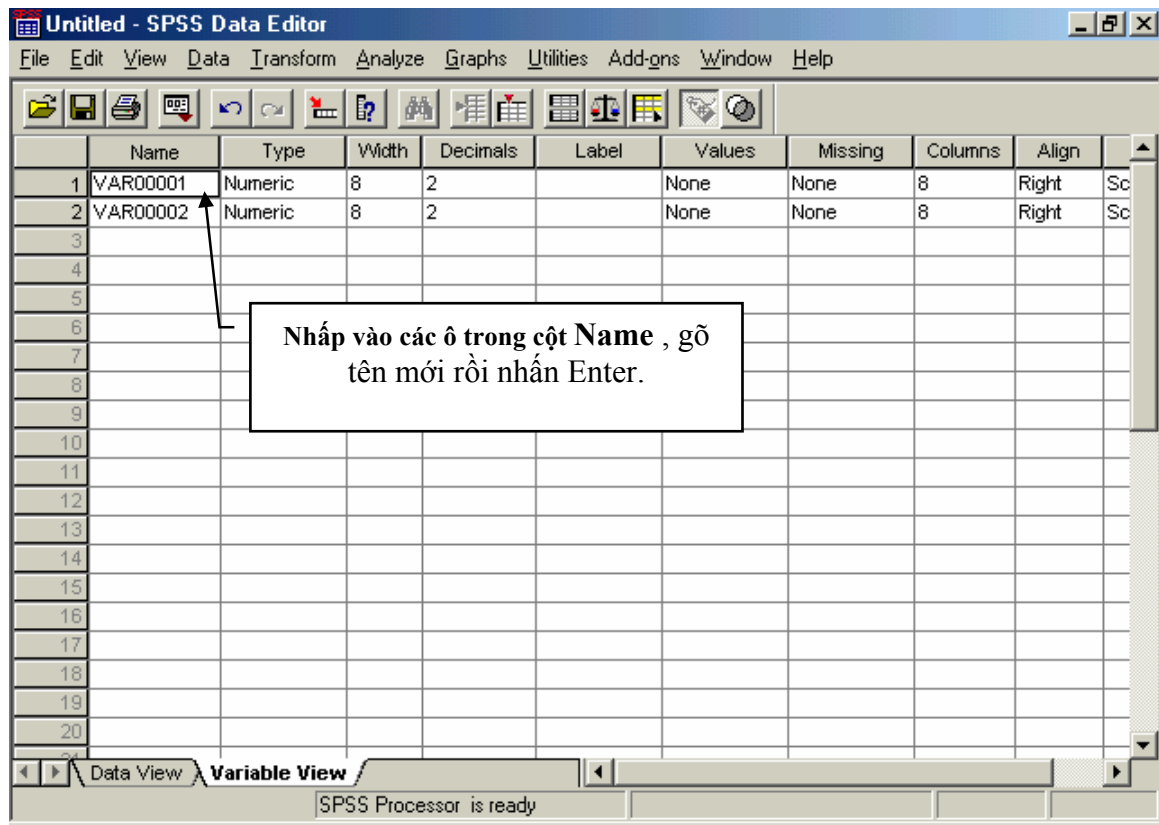
Do Copy trực tiếp nên dòng đầu tiên chứa tên biến, không phải là giá trị nên SPSS bỏ trống dòng đó. Mặc khác, tên biến được mặc định là **VAR0001** và **VAR0002**.

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

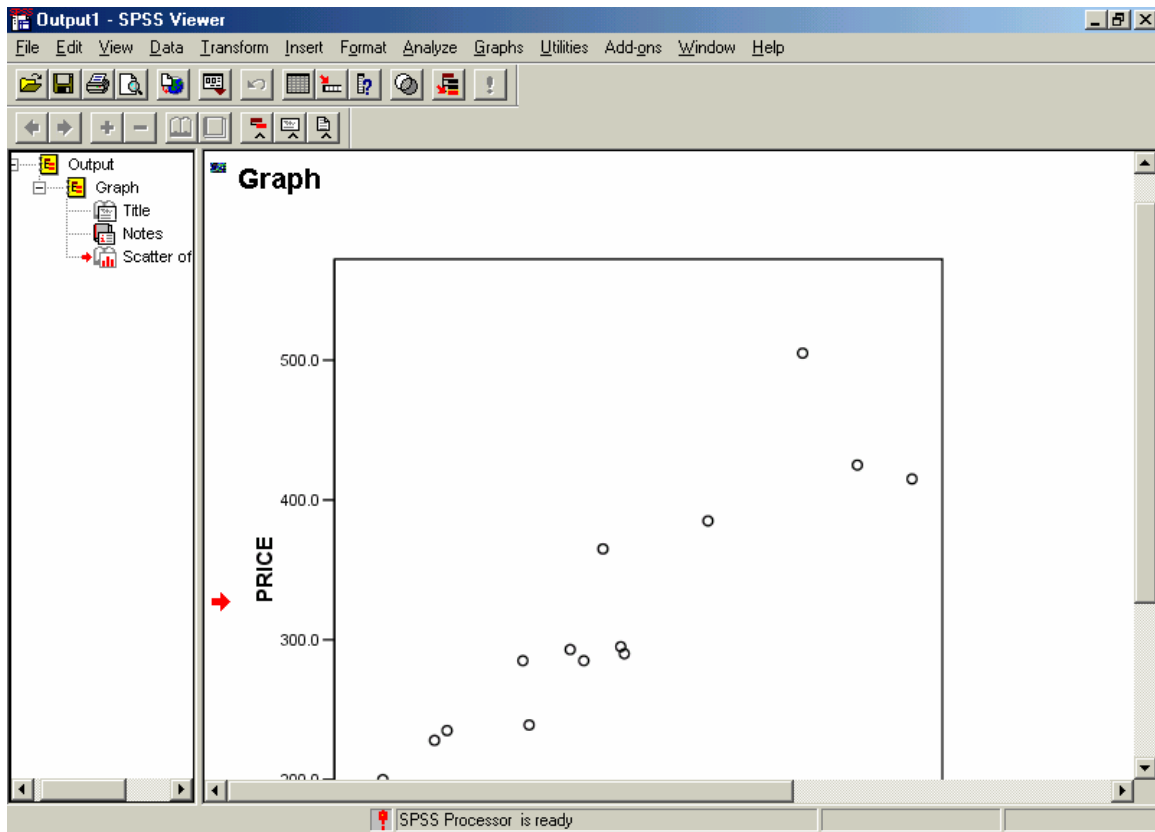
Nhập vào biến **PRICE** là biến phụ thuộc.

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.

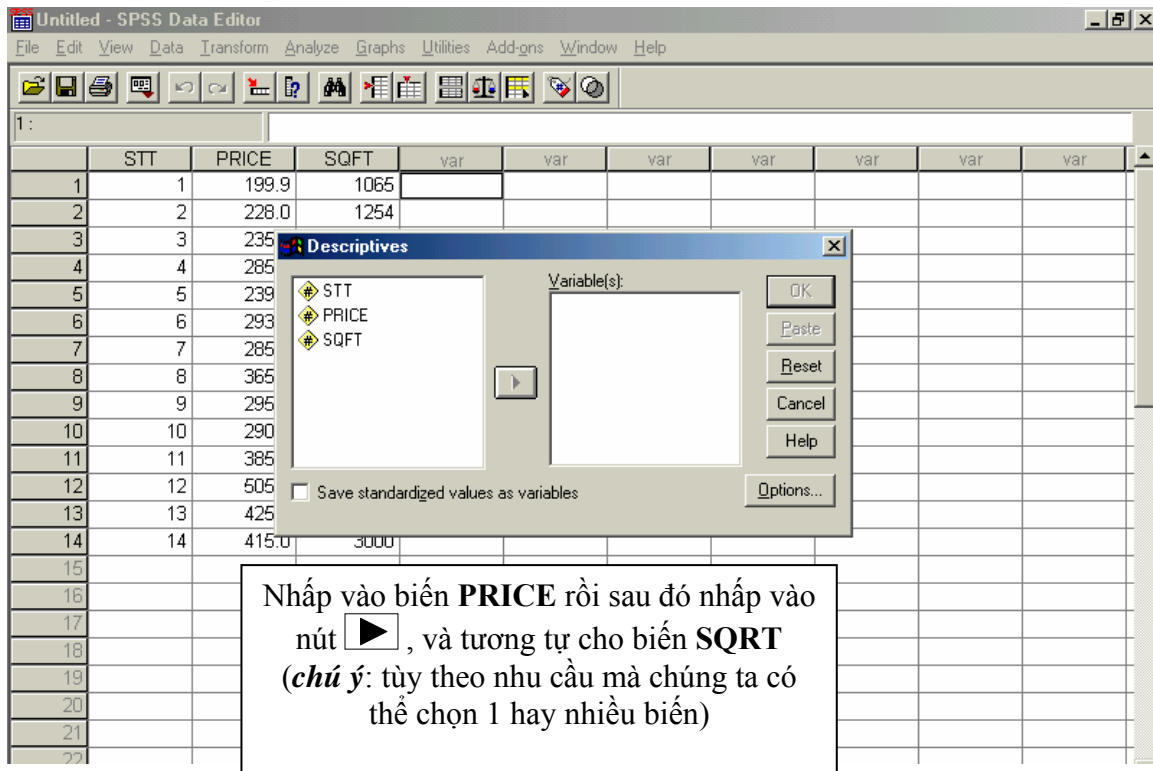
HÌNH 18



6. Thống kê mô tả dữ liệu

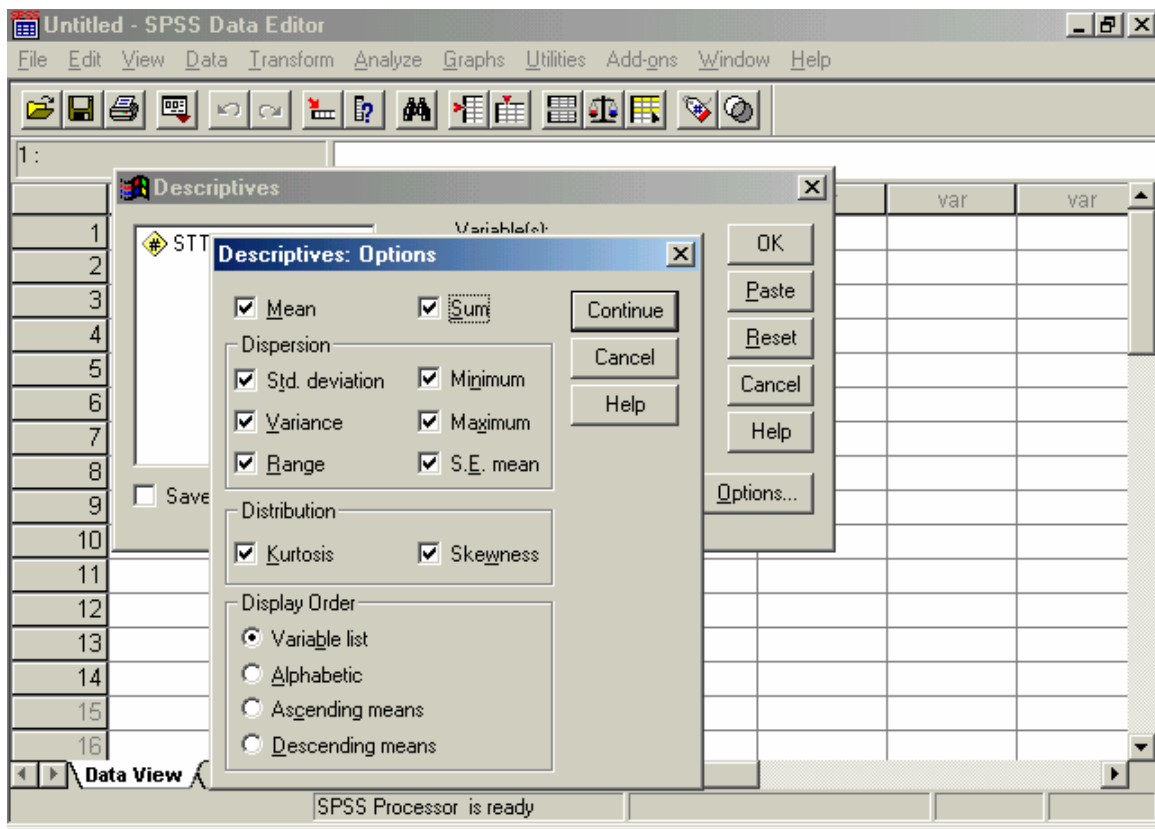
Vào Menu **Analyze, Descriptive Statistics, Descriptives...**

HÌNH 19



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

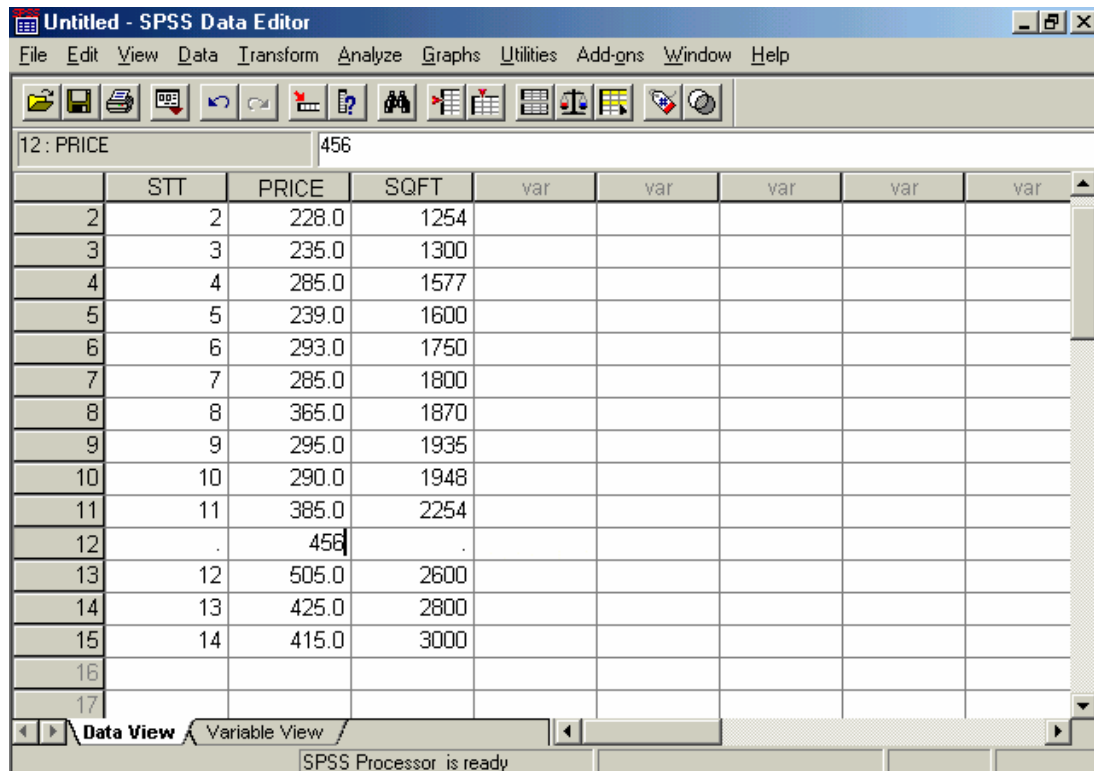
HÌNH 22

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 interface. The title bar reads '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 columns: STT, PRICE, SQFT, and five columns labeled 'var'. The data rows are numbered 2 through 17. Row 12 is highlighted, and the value '456' is entered in the PRICE column. The status bar at the bottom indicates 'SPSS Processor is ready'.

	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								

b. Thêm biến

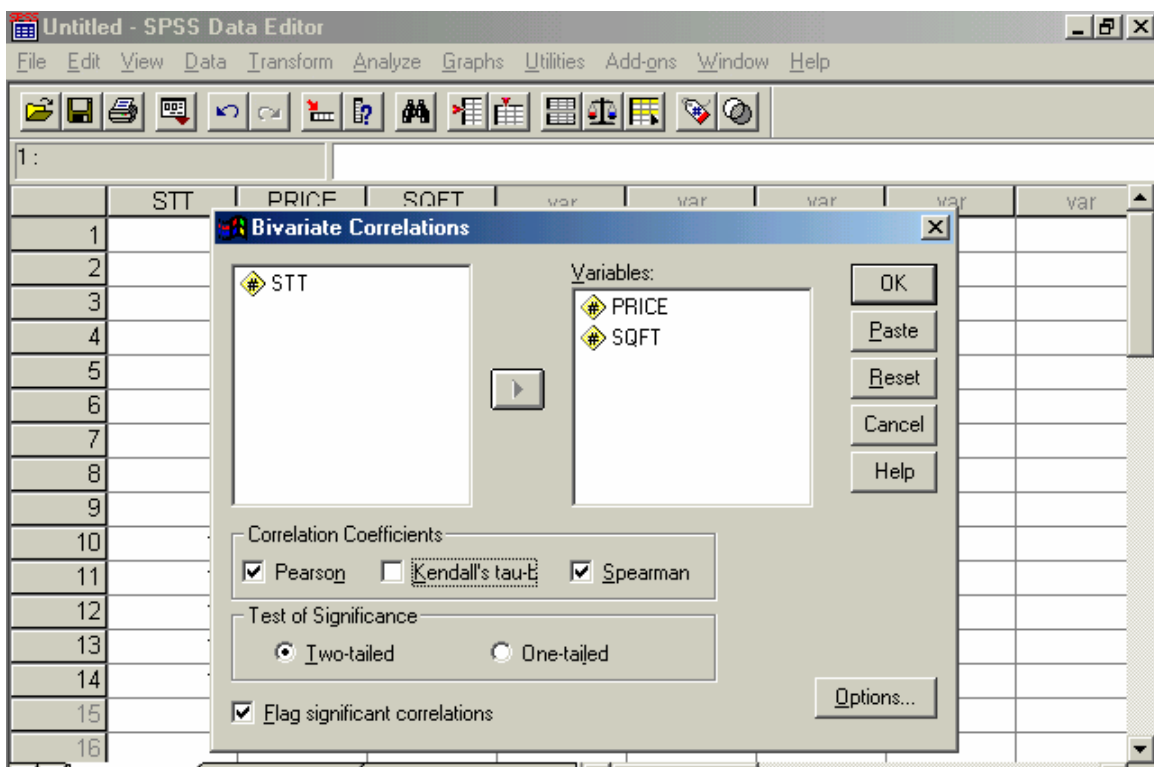
Thao tác để thêm biến cũng giống như trong Excel, chúng ta có thể gõ trực tiếp vào trong các ô của biến mới. Sau đó vào **Variable View** để đặt tên cho biến mới đó.

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**.

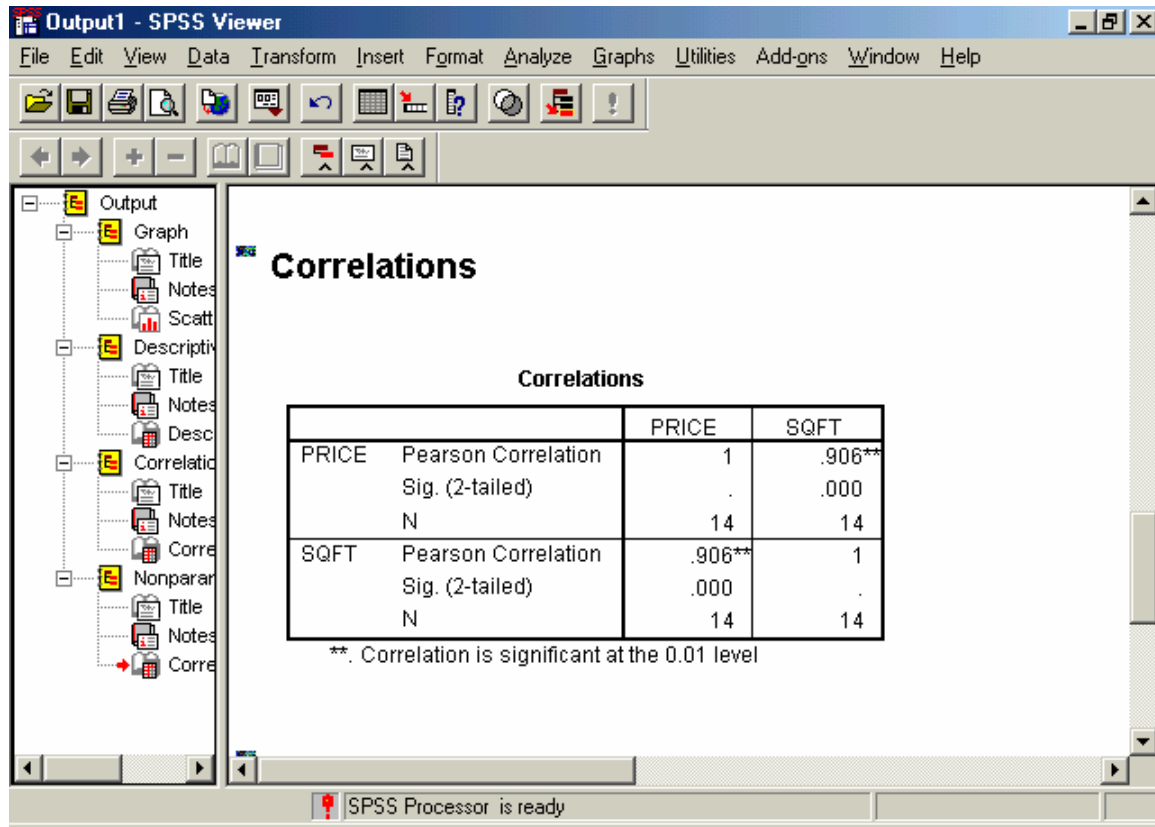
HÌNH 24



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

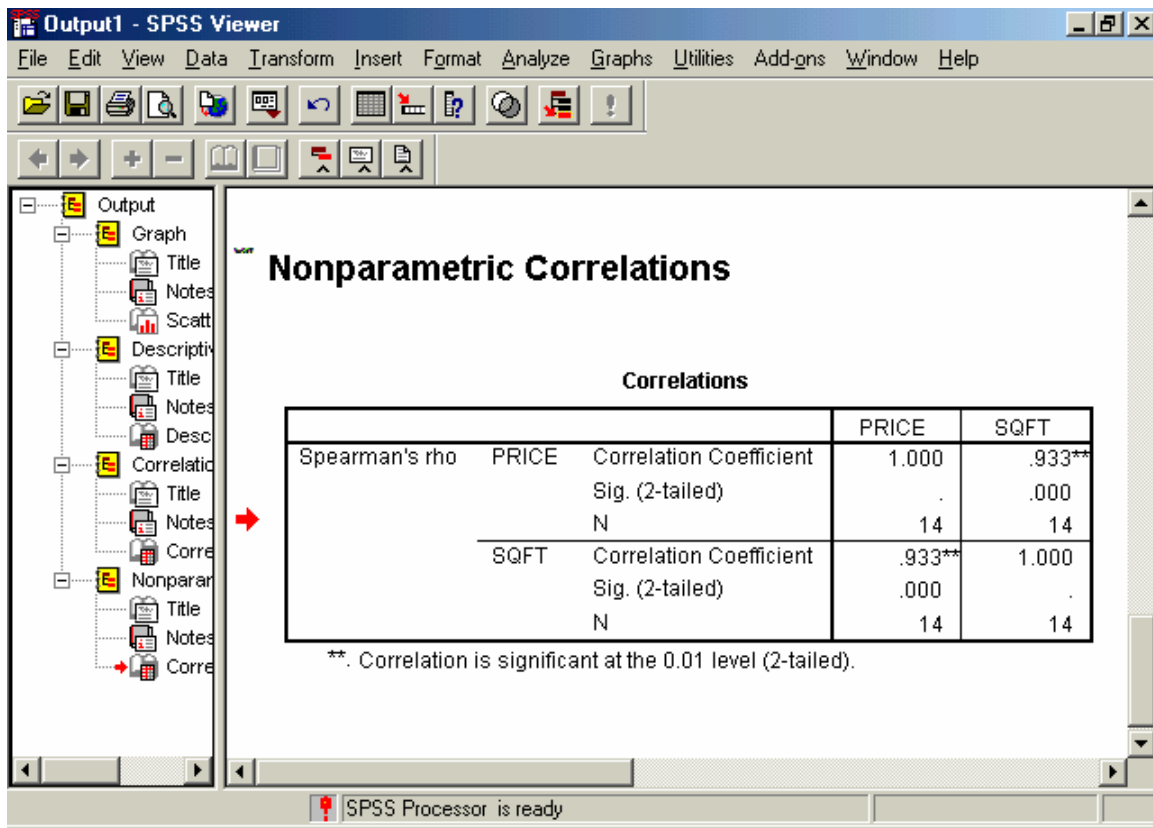
Đây là bảng kết quả của hệ số **Pearson**.

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:

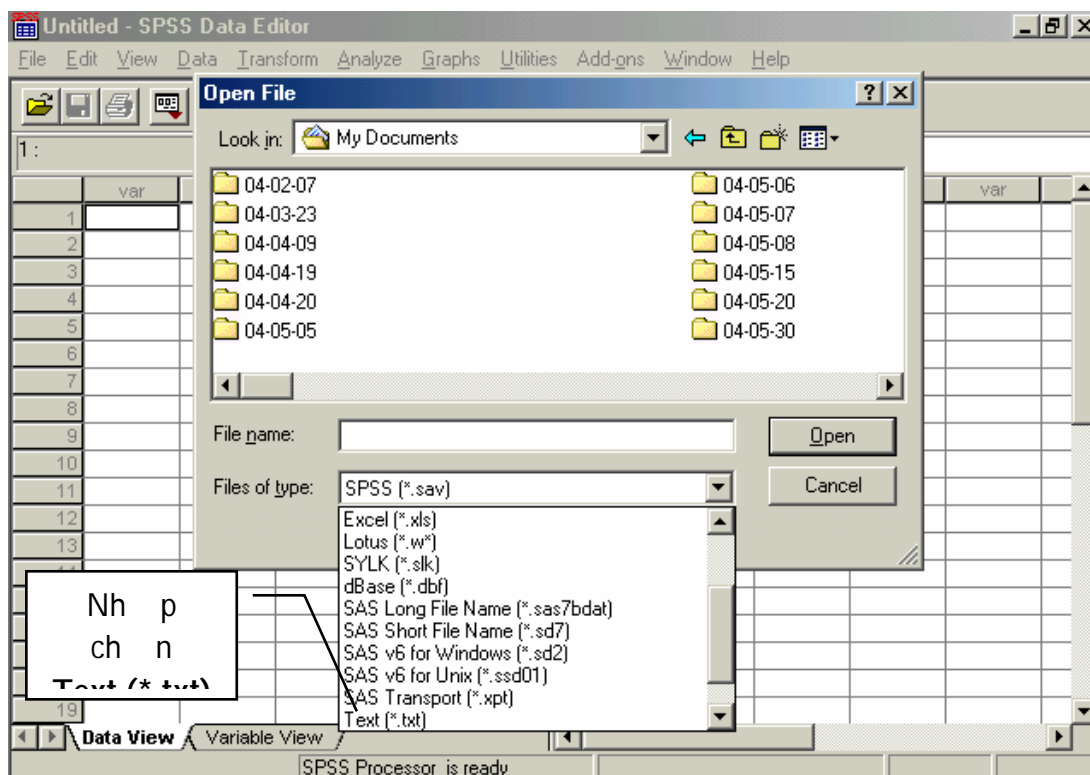
1. Mở dữ liệu từ tập tin dạng text
2. Xem xét sự tương quan giữa các biến
3. Hồi quy OLS (trường hợp đơn biến)
4. Hồi quy OLS (trường hợp đa biến)
5. Hồi quy trong trường hợp có hiện tượng đa cộng tuyến hoàn hảo

* 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).

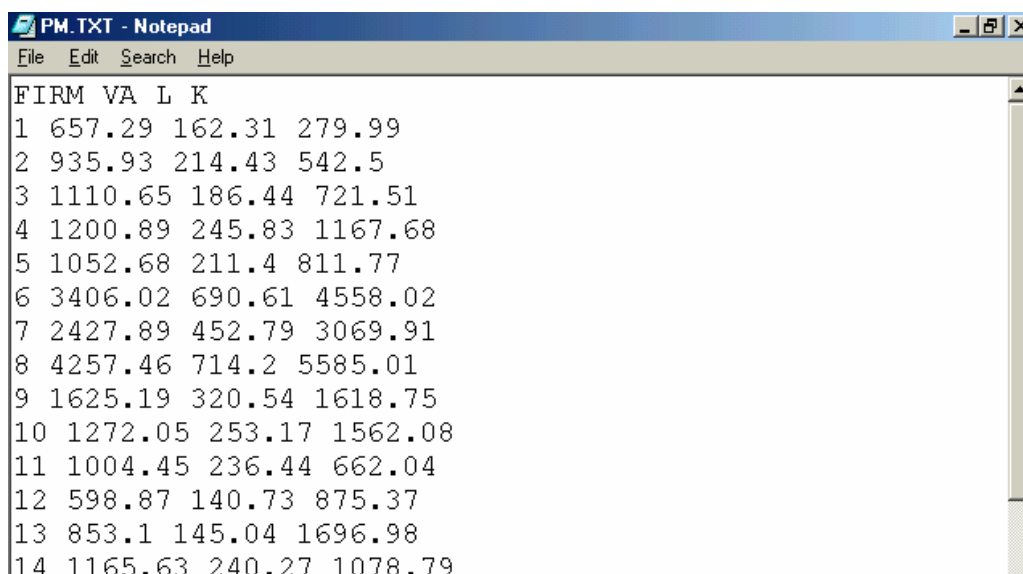
HÌNH 1



Ví dụ:

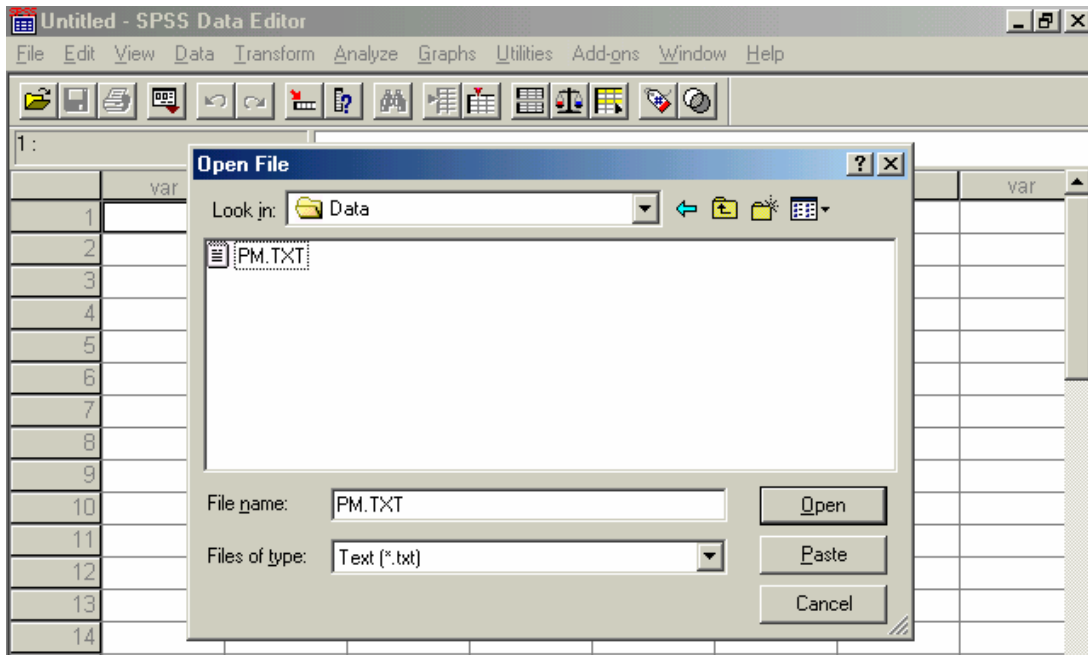
Chúng ta sẽ mở tập tin **pm.txt** chứa dữ liệu về **FIRM, VA, K** và **L** đã thực hành trong **EVIEWS**.

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**

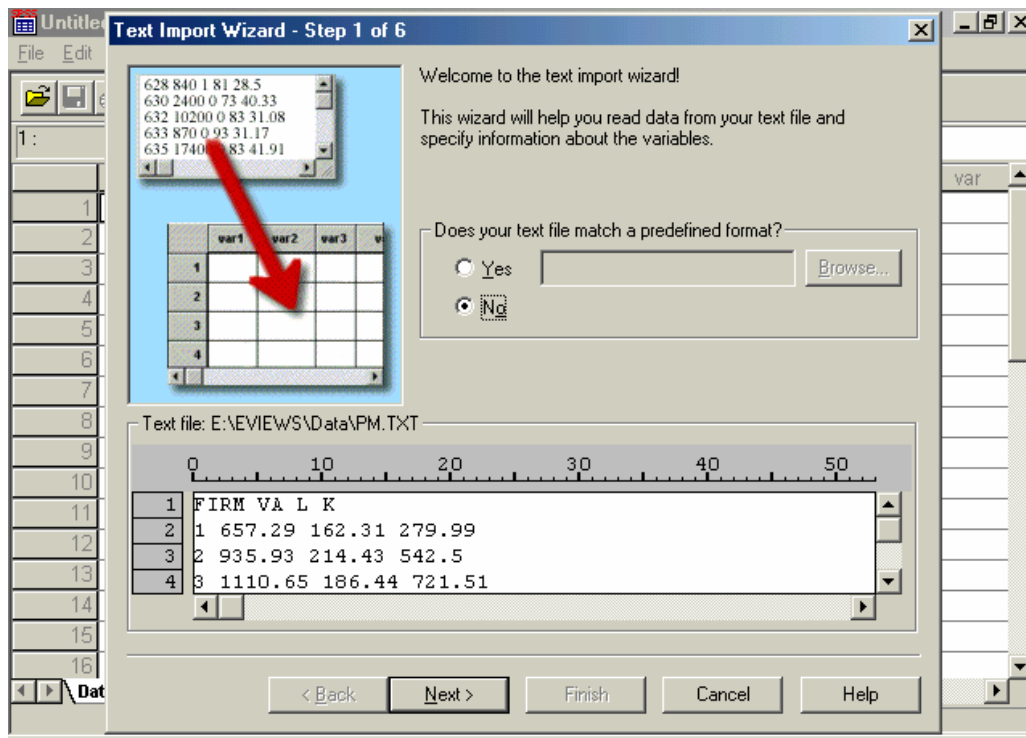
HÌNH 3



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

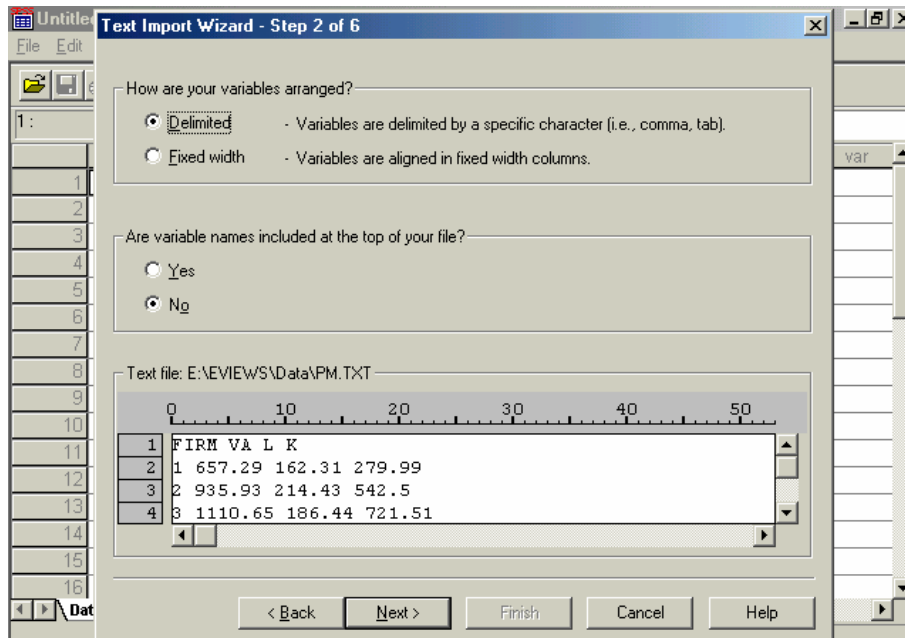
- **Bước 1 (Hình 4):** Trong Bước 1, SPSS mặc định sẵn ở chế độ **No**, chúng ta chỉ việc nhấp **Next**.

HÌNH 4



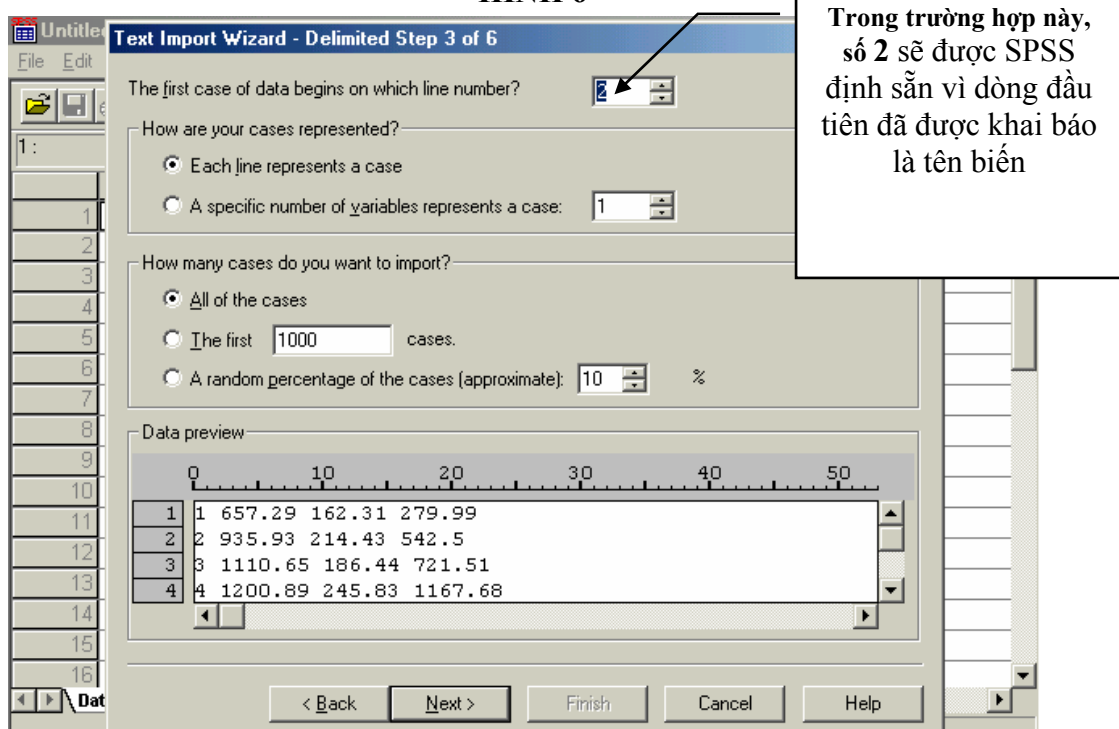
- **Bước 2 (Hình 5):** Trong Bước 2, cần chú ý dòng chữ **Are variable names included at the top of your file?**, ngụ ý hỏi có phải tên biến nằm ở dòng đầu tiên của tập tin đó không? Nếu phải thì nhấp **Yes**, ngược lại thì **No**. Trong trường hợp của tập tin **pm.txt** thì dòng đầu tiên có chứa tên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

HÌNH 5



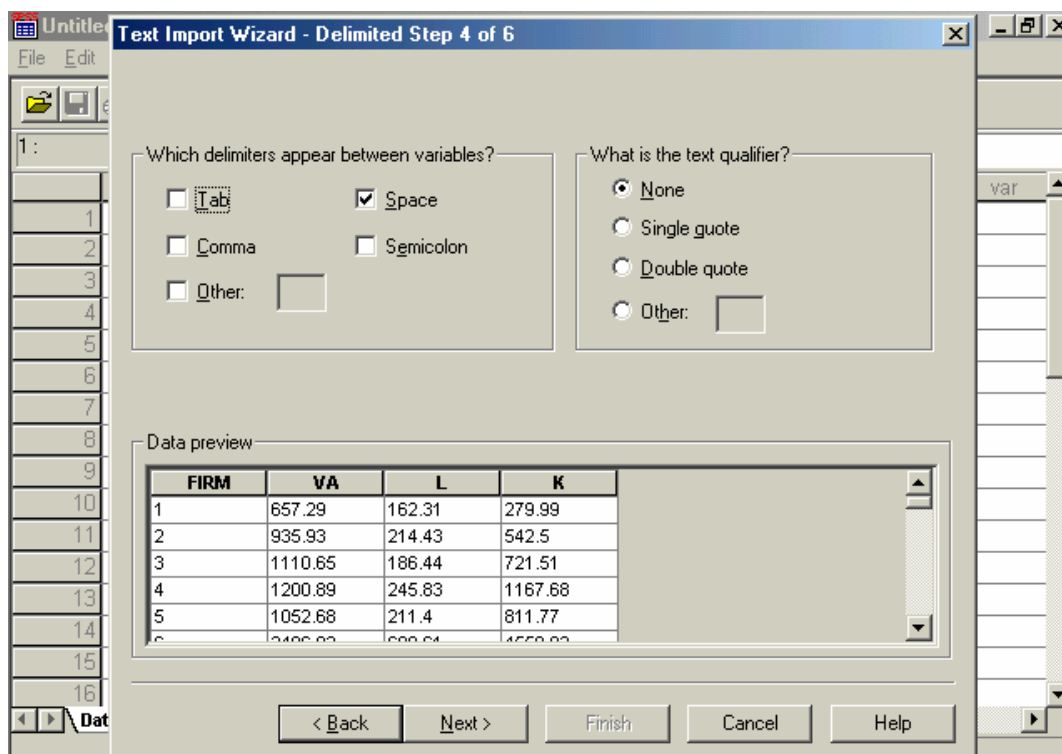
- **Bước 3 (Hình 6):** Chú ý dòng chữ **The first case of data begins on which line number?**, ngụ ý hỏi dữ liệu sẽ bắt đầu từ dòng thứ mấy. Khai báo xong, bấm **Next**.

HÌNH 6



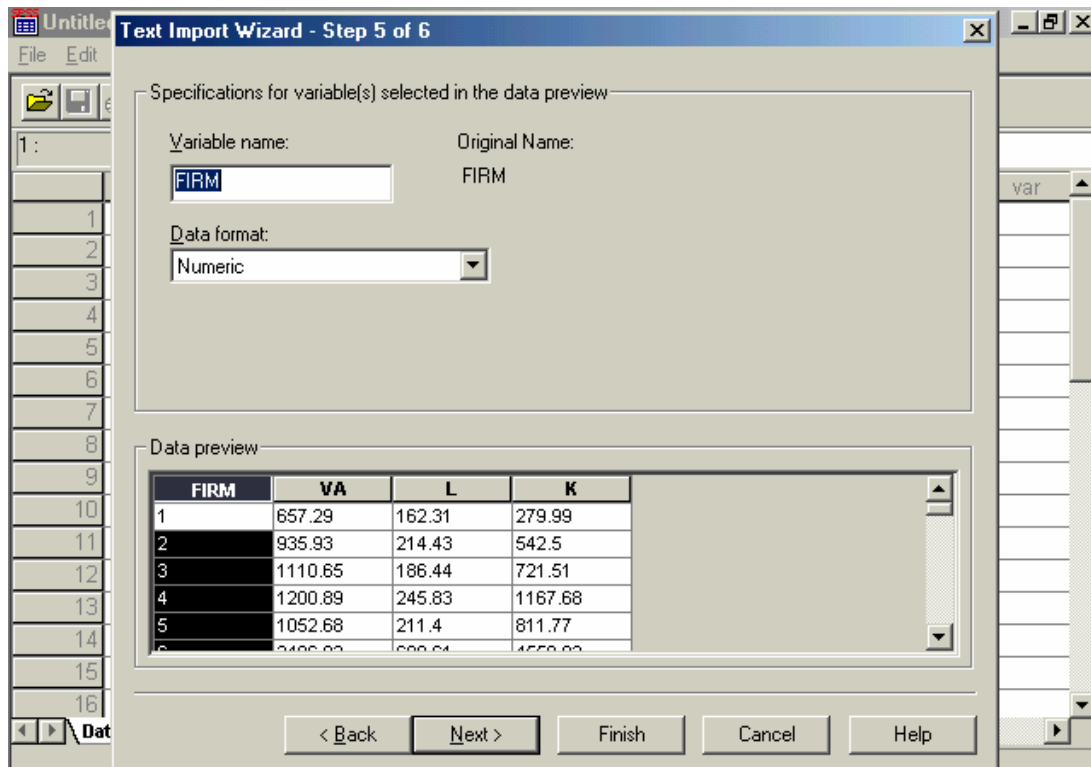
- **Bước 4 (Hình 7):** Chú ý đến câu **Which delimiters appear between variables?**, ngụ ý hỏi dữ liệu của các biến được phân định bằng cách nào. Bằng **Tab**, bằng khoảng trống (**Space**), bằng dấu phẩy (**Comma**), dấu chấm phẩy (**Semicolon**) hay ở dạng cụ thể nào khác (**Other**). Ở đây SPSS thường 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**.

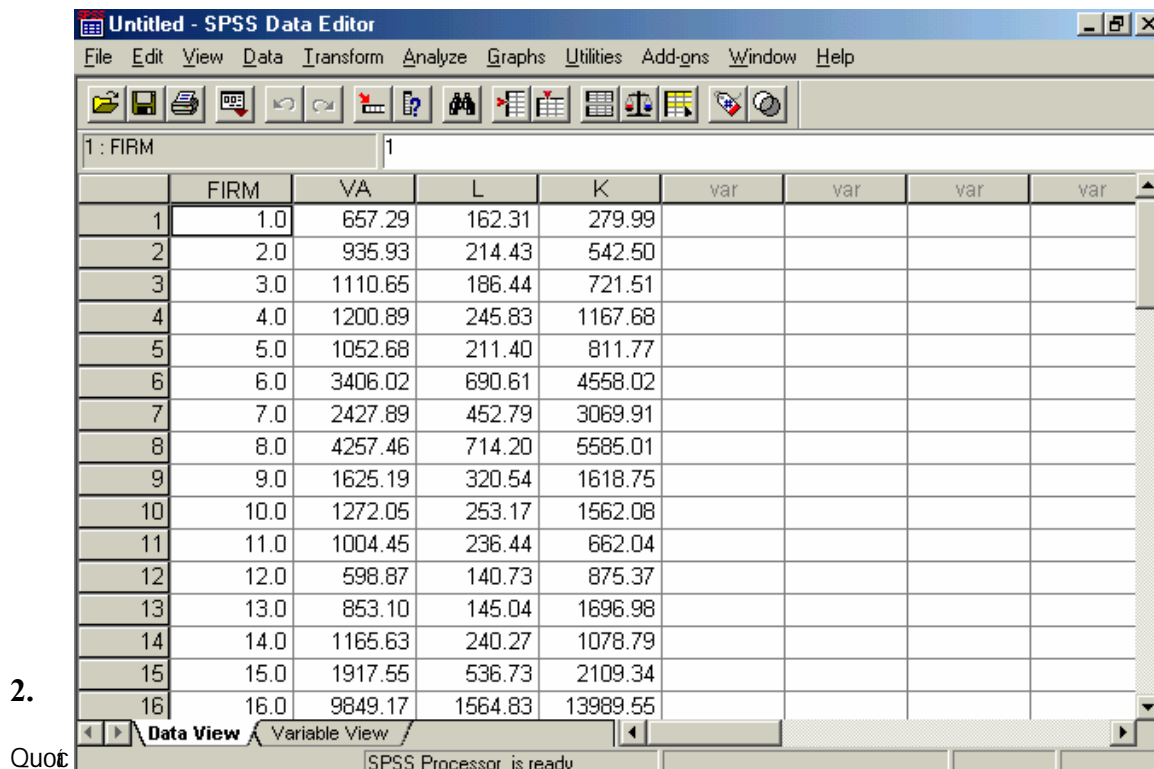
HÌNH 8



- **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

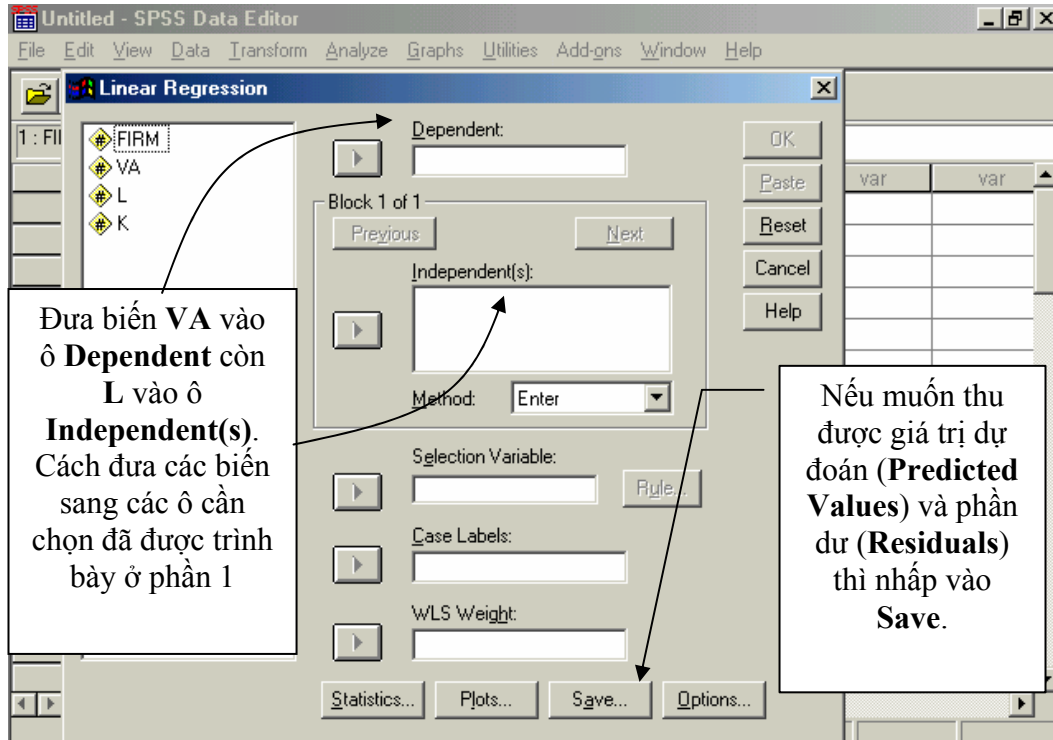
		VA	L	K
VA	Pearson Correlation	1	.965**	.975**
	Sig. (2-tailed)	.	.000	.000
	N	27	27	27
L	Pearson Correlation	.965**	1	.964**
	Sig. (2-tailed)	.000	.	.000
	N	27	27	27
K	Pearson Correlation	.975**	.964**	1
	Sig. (2-tailed)	.000	.000	.
	N	27	27	27

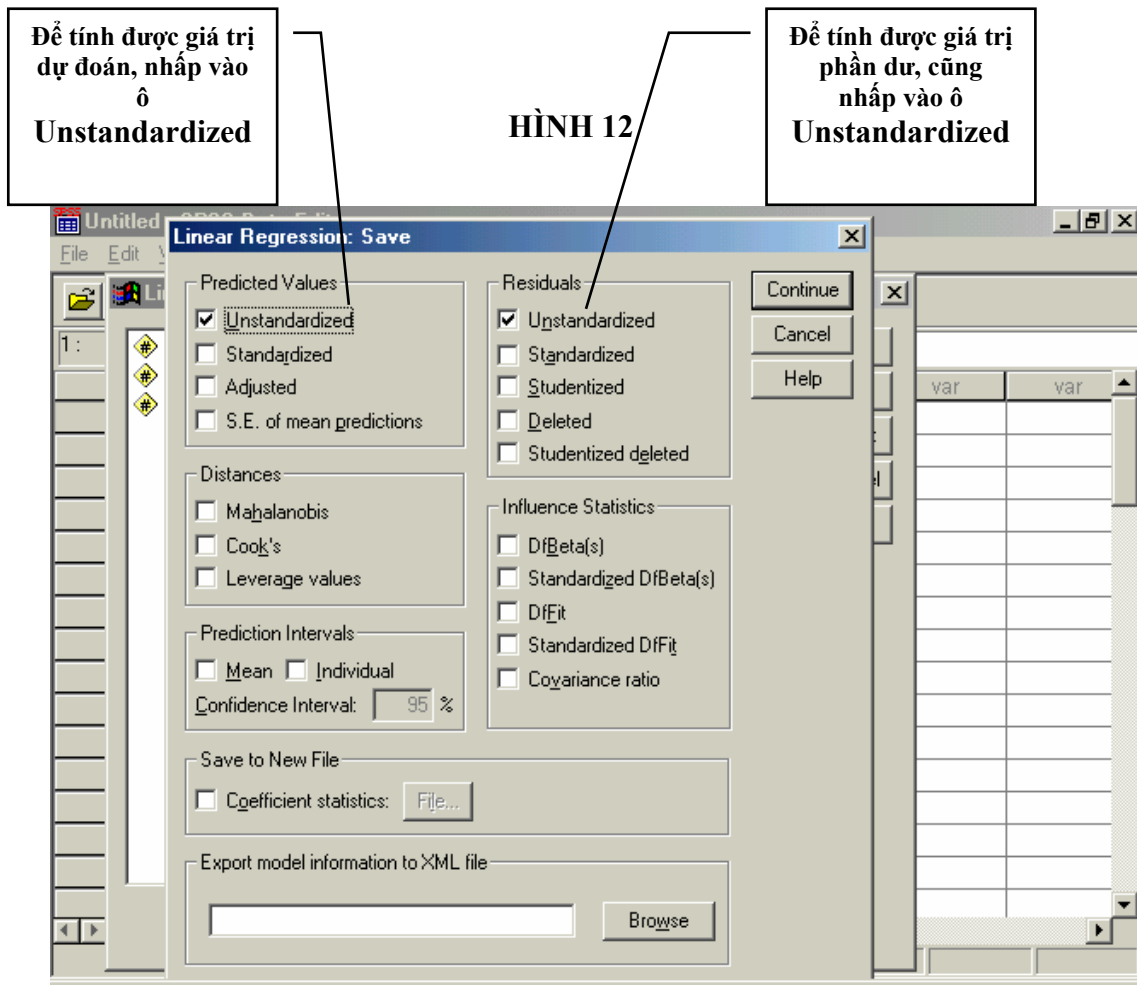
** . Correlation is significant at the 0.01 level (2-tailed).

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.

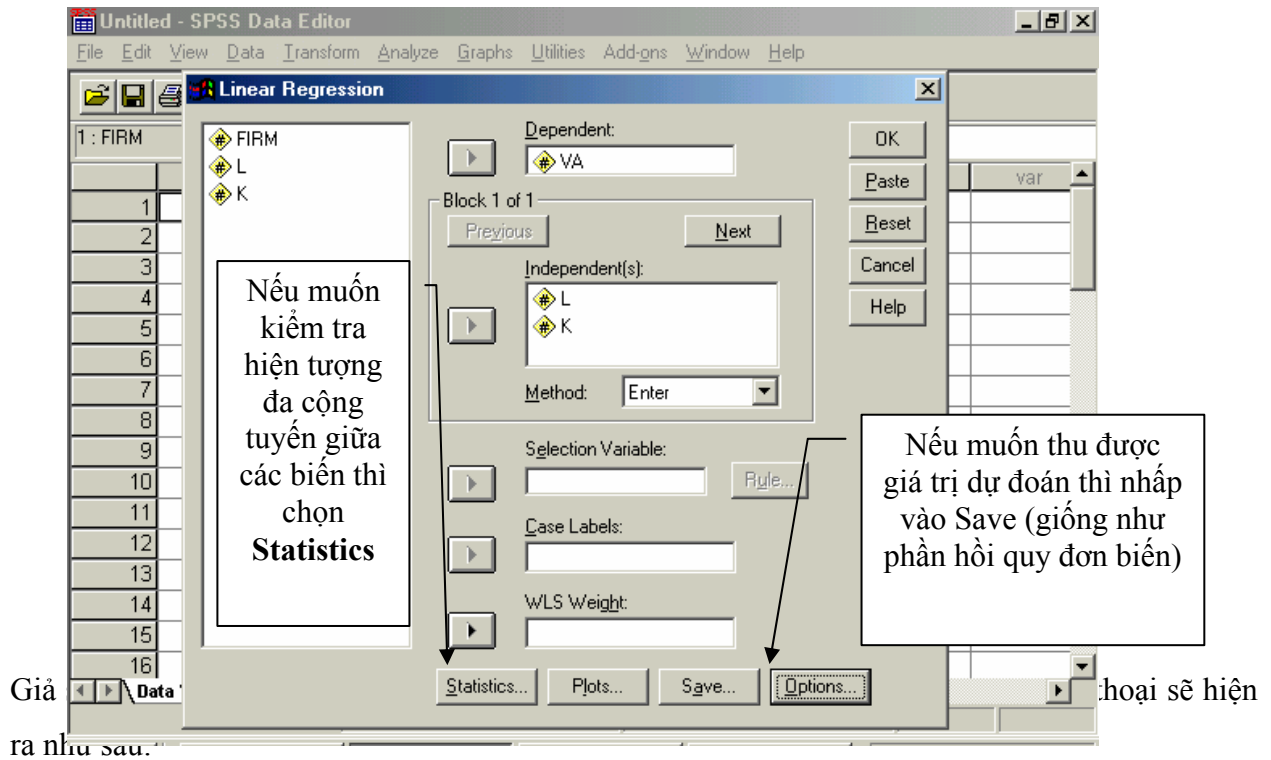
HÌNH 13

	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	

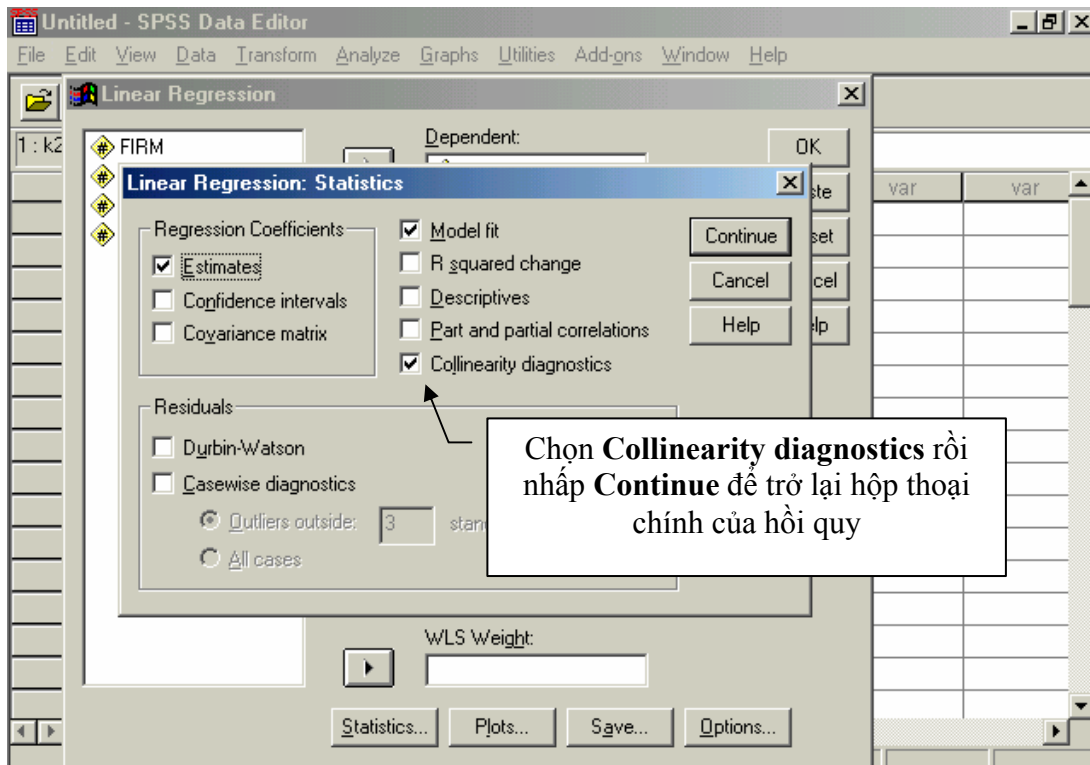
4. Hồi quy OLS (trường hợp đa biến)

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:

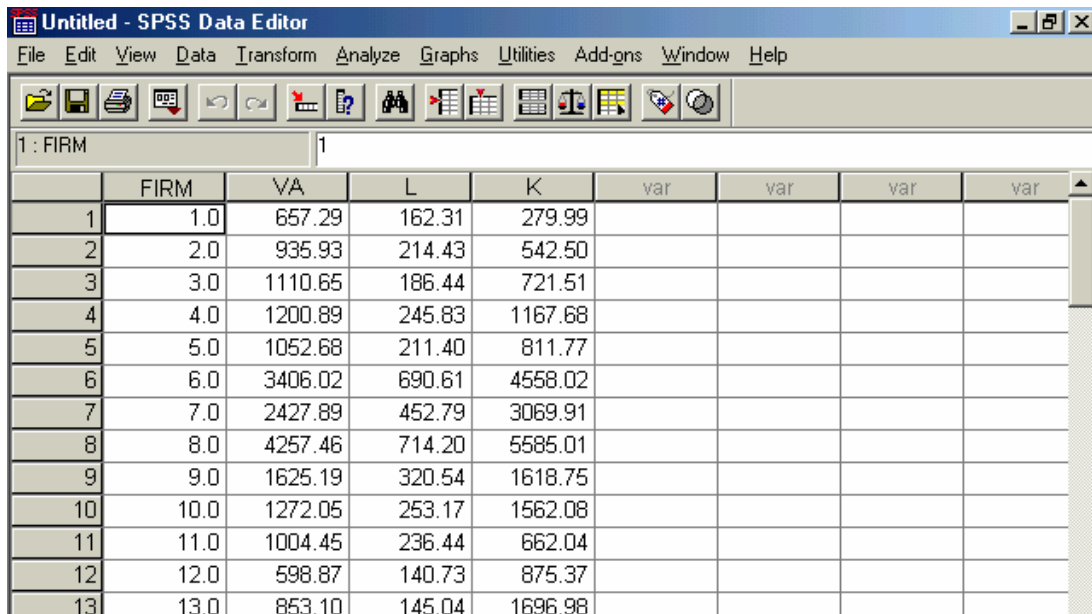
- ❖ Tạo biến mới từ các biến đã có trong dữ liệu

* 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

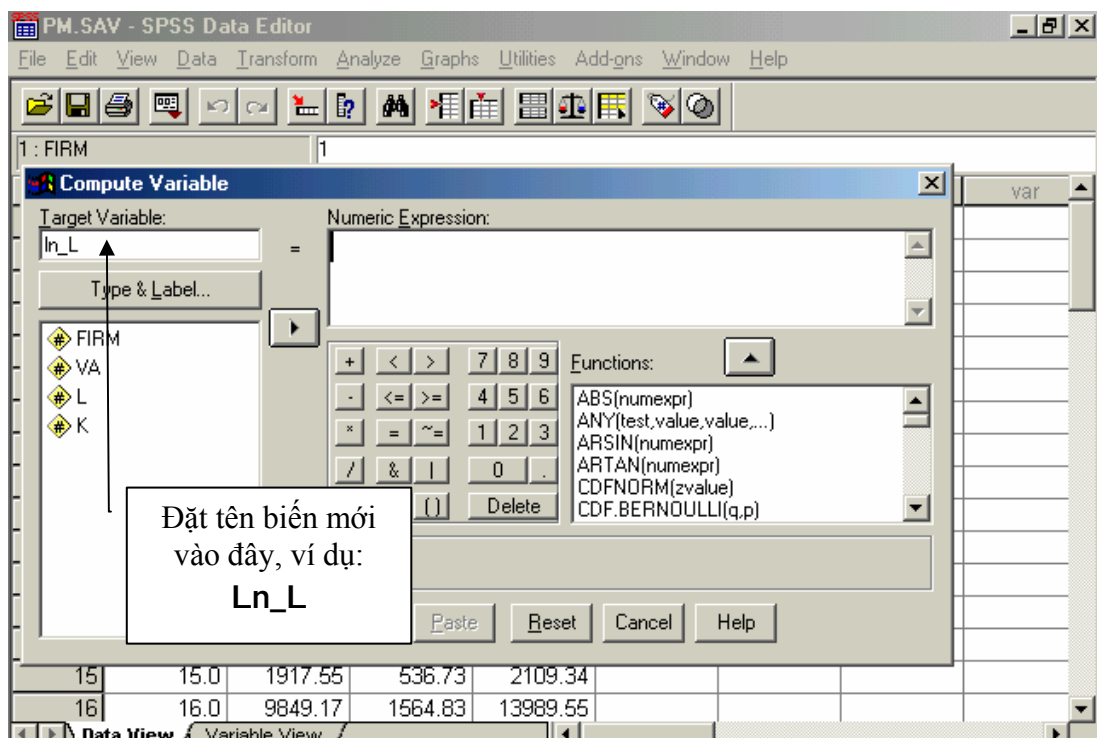


	FIRM	VA	L	K	var	var	var	var
1	1.0	657.29	162.31	279.99				
2	2.0	936.93	214.43	542.50				
3	3.0	1110.65	186.44	721.51				
4	4.0	1200.89	245.83	1167.68				
5	5.0	1052.68	211.40	811.77				
6	6.0	3406.02	690.61	4558.02				
7	7.0	2427.89	452.79	3069.91				
8	8.0	4257.46	714.20	5585.01				
9	9.0	1625.19	320.54	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				

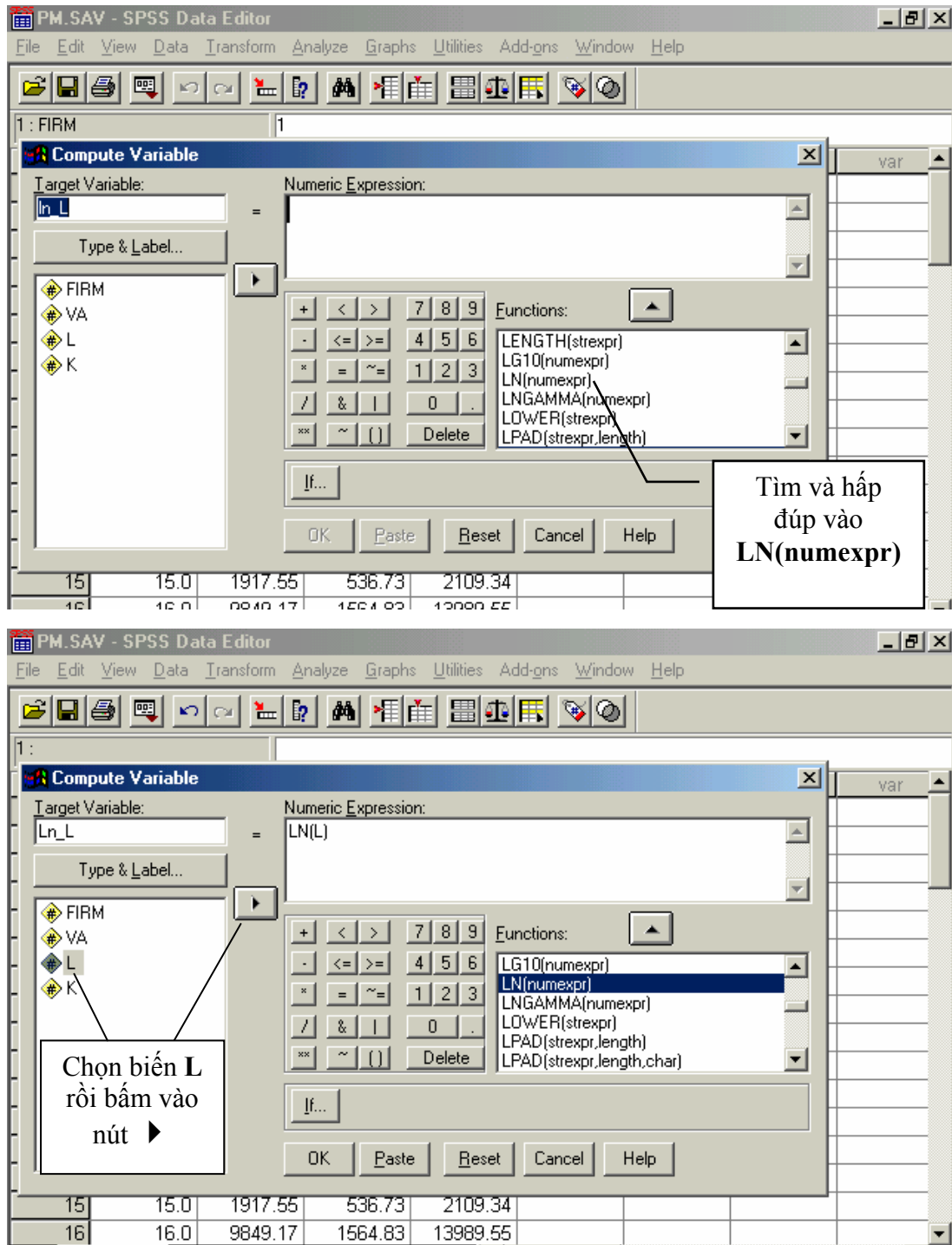
- a. Giả sử, chúng ta cần hồi quy VA theo $\ln(L)$ và $\ln(K)$, như vậy cần tạo thêm 2 biến mới là \ln_L và \ln_K từ L và K.

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

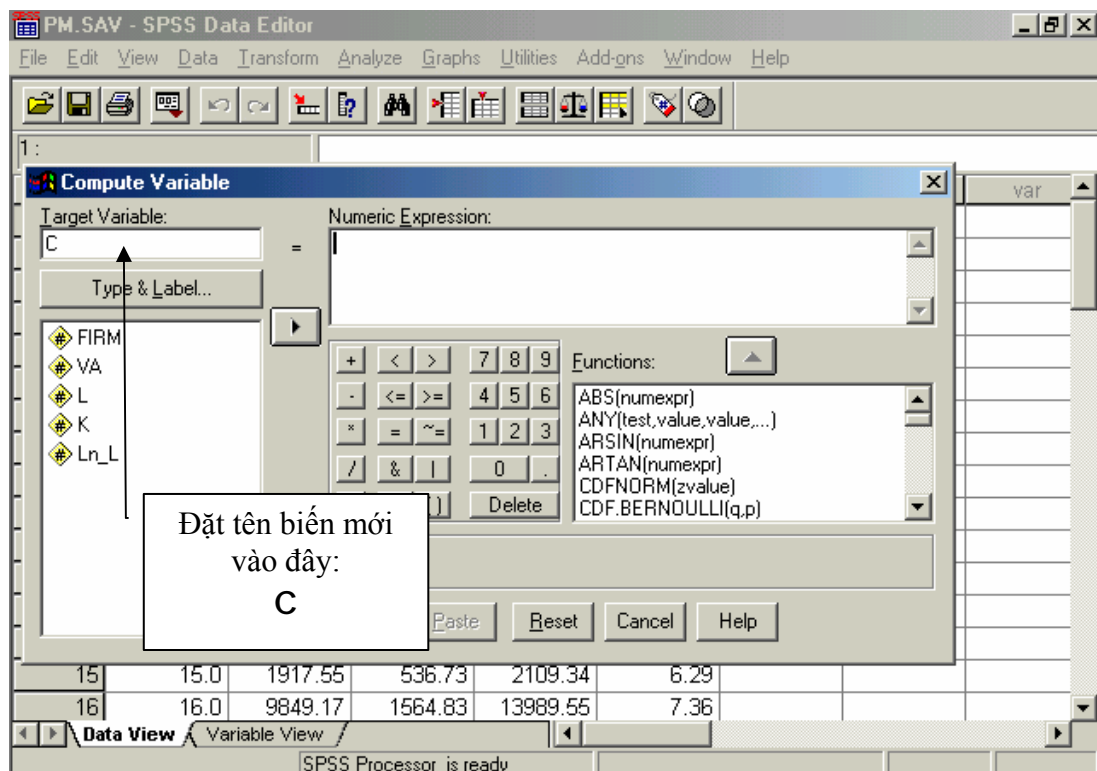
	FIRM	VA	L	K	Ln_L	var	var	var
1	1.0	657.29	162.31	279.99	5.09			
2	2.0	935.93	214.43	542.50	5.37			
3	3.0	1110.65	186.44	721.51	5.23			
4	4.0	1200.89	245.83	1167.68	5.50			
5	5.0	1052.68	211.40	811.77	5.35			
6	6.0	3406.02	690.61	4558.02	6.54			
7	7.0	2427.89	452.79	3069.91	6.12			
8	8.0	4257.46	714.20	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 image shows two screenshots from the SPSS Data Editor. The top screenshot displays the 'Compute Variable' dialog box. The 'Target Variable' is 'C' and the 'Numeric Expression' is 'K'. A text box with an arrow pointing to the 'K' variable in the list on the left says 'Chọn biến K rồi bấm vào nút ►'. Another text box with an arrow pointing to the 'OK' button says 'Sau đó tiếp tục gõ vào *2+1, rồi bấm OK'. The bottom screenshot shows the data table with the new variable 'C' added, containing values calculated from 'K' multiplied by 2 plus 1. A text box on the right says 'Trong dữ liệu đã xuất hiện thêm biến C'.

	FIRM	VA	L	K	Ln L	C	var	var
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

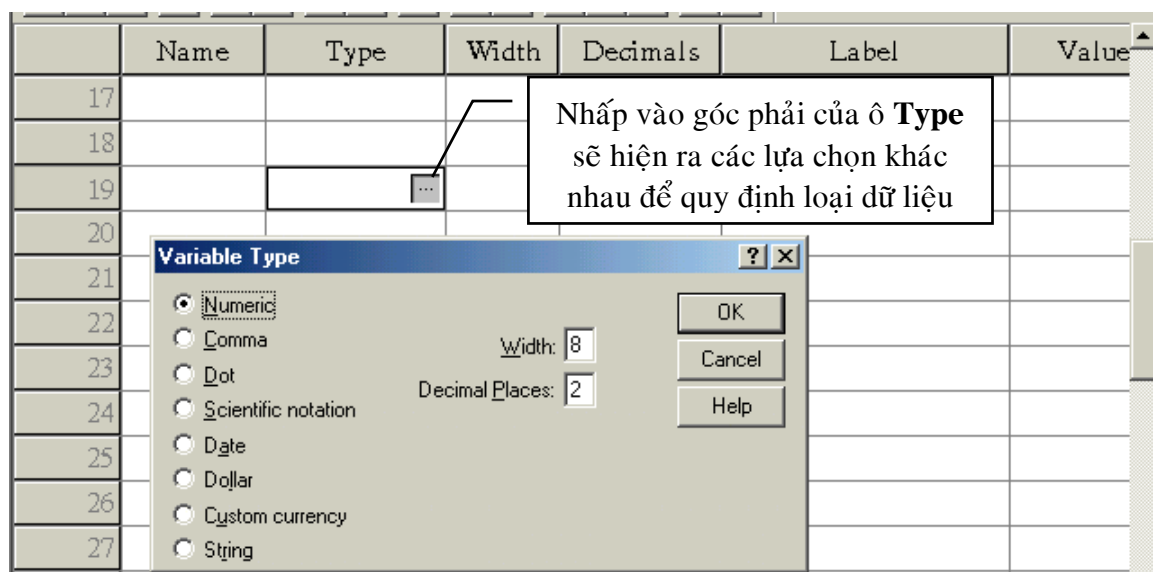
Khi sử dụng SPSS 12.0, chúng ta thường thấy hai Sheet: **Data View** và **Variable View**. **Data View** chứa dữ liệu còn **Variable View** chứa các thông tin của các biến trong dữ liệu. Các thông tin này bao gồm:

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

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

- a. **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 “.”)
- b. **Type:** loại dữ liệu của biến.

HÌNH 1



c. **Measure:** thang đo

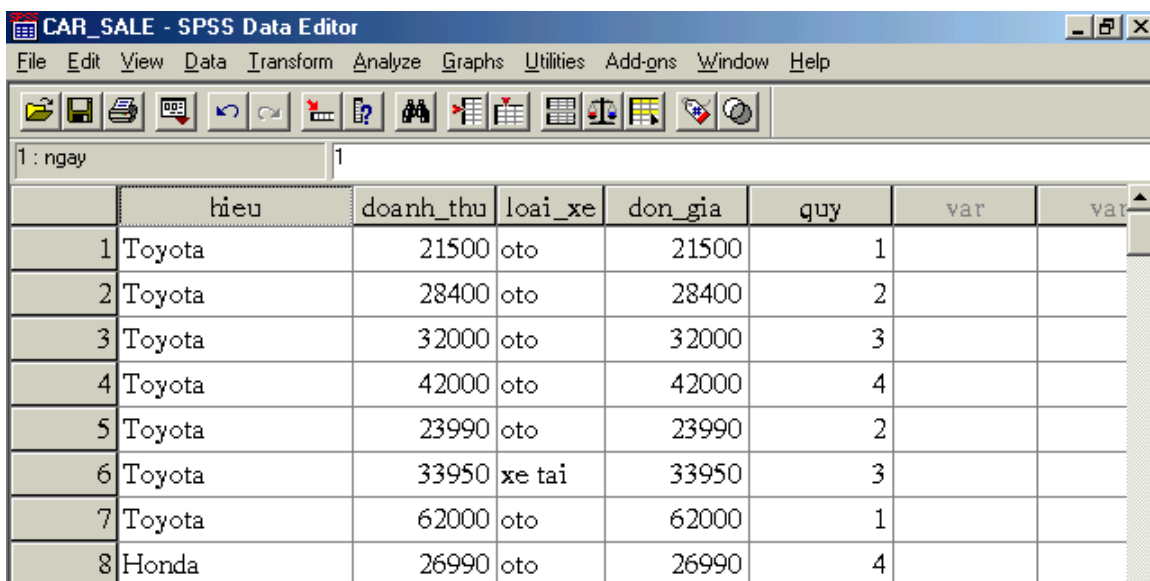
Trong SPSS 12.0 có 3 loại thang đo: Scale, Nominal và Ordinal.

- **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



	hiệu	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.

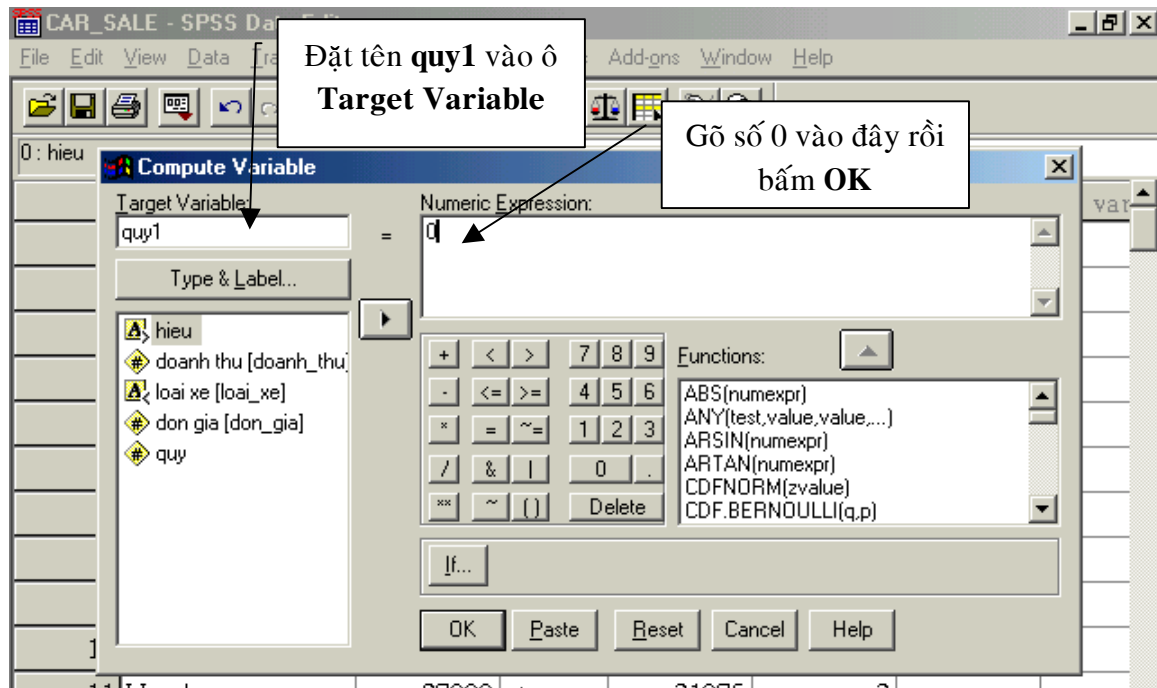
Trong đó:

- hiệu: 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**

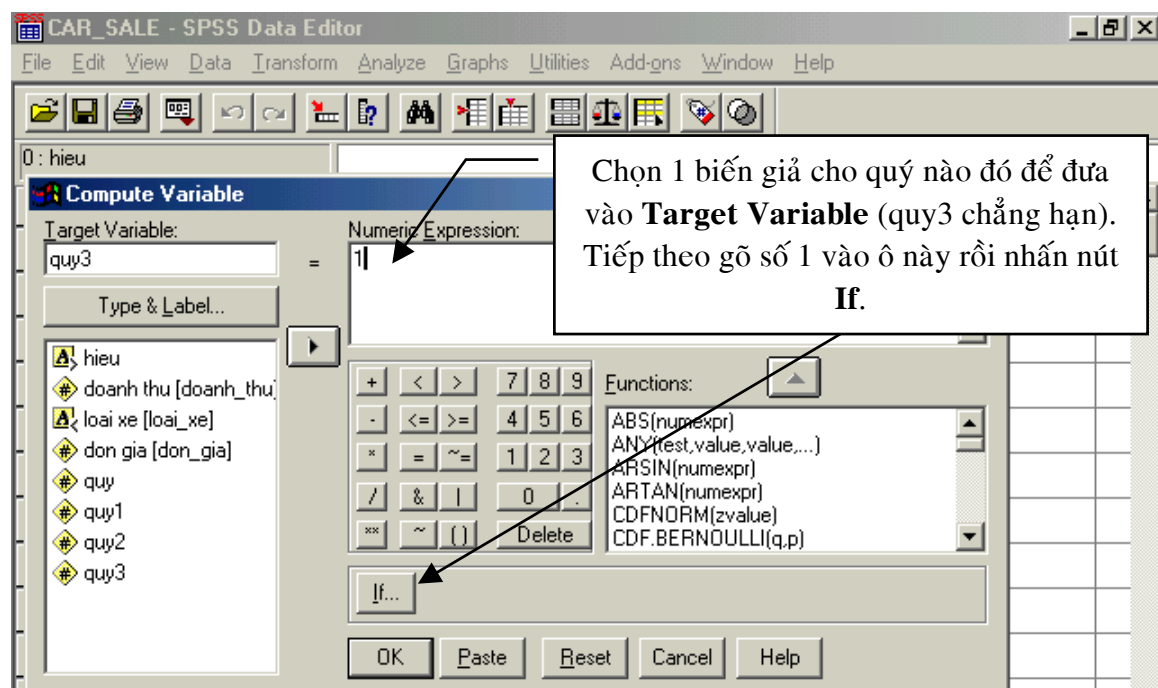
HÌNH 3



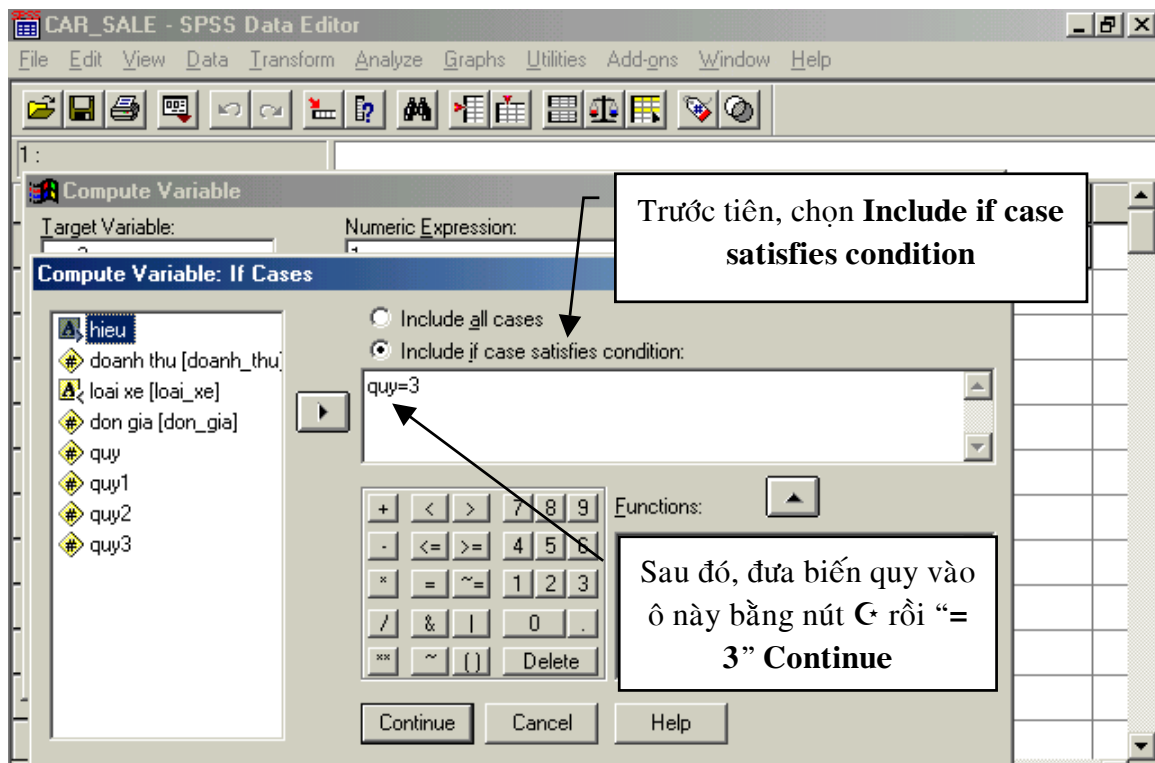
Lần lượt thực hiện như thế cho **quy2** và **quy3**. Lúc này trong dữ liệu đã xuất hiện biến **quy1**, **quy2** và **quy3** với tất cả các giá trị đều bằng 0.

Tiếp theo, lại trở vào **Compute**.

HÌNH 4



HÌNH 5

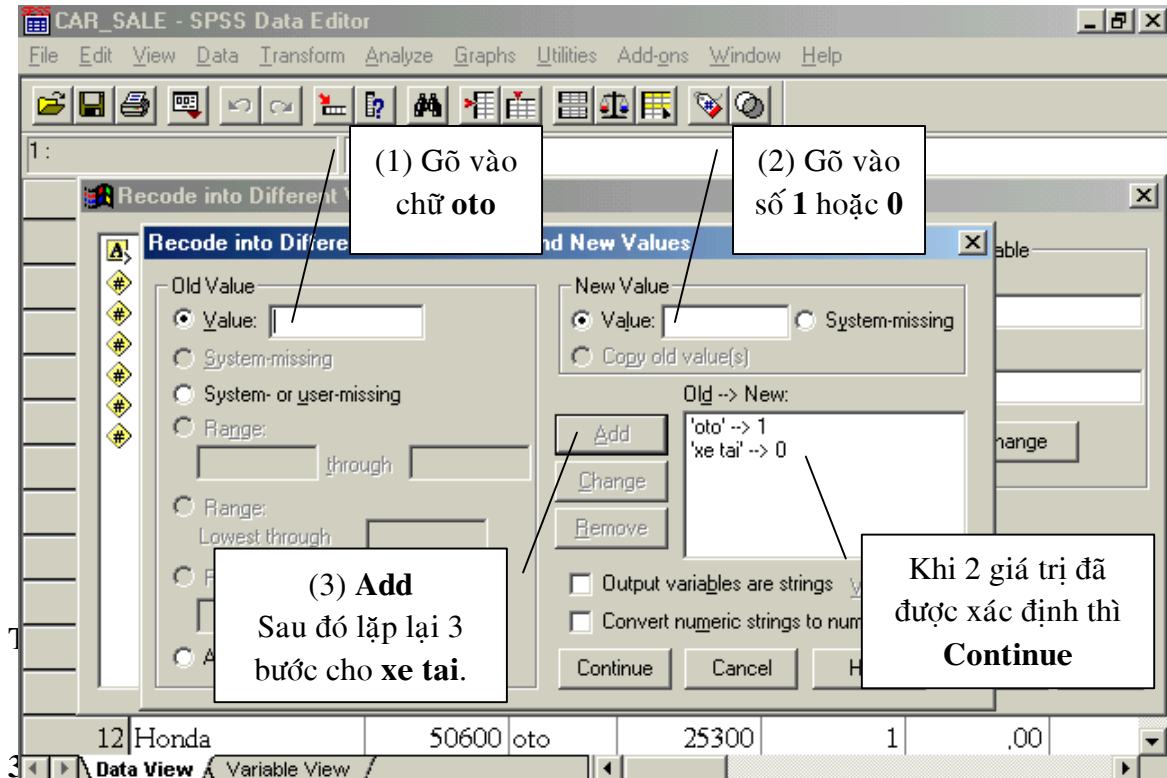
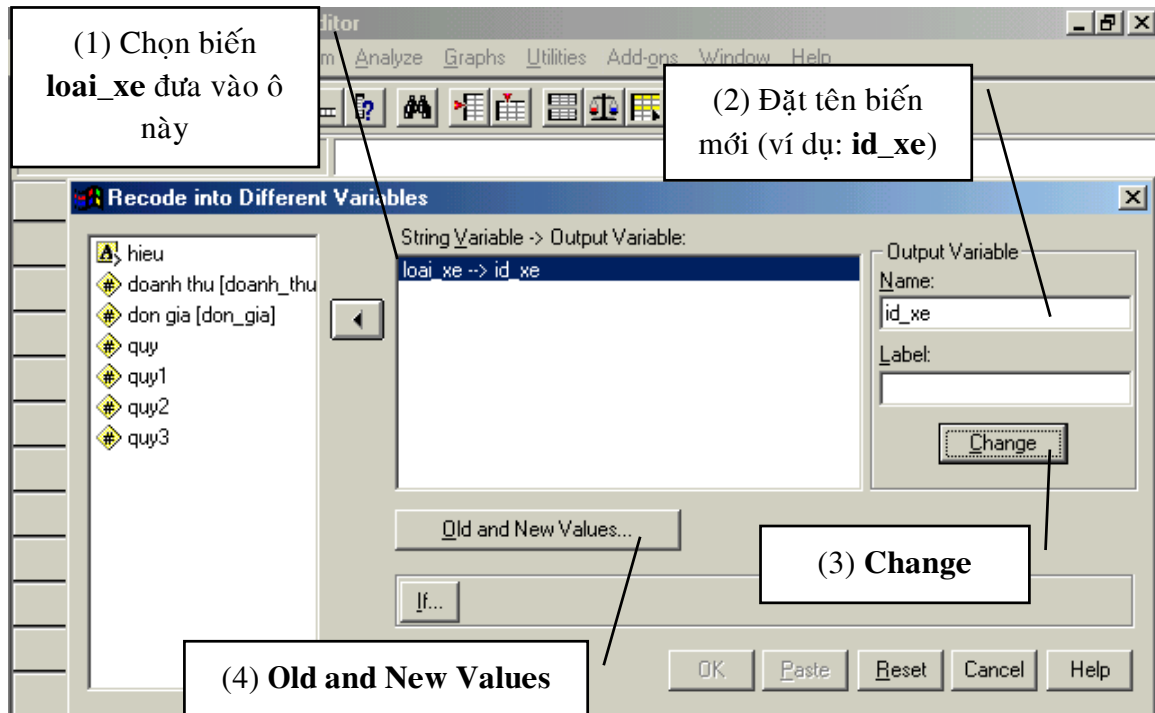


Trở lại hộp thoại phía trước, tiếp tục chọn OK để hoàn tất tạo biến giả **quy3**, và tiếp tục làm tương tự cho quy1 và quy2.

2.2. Tạo biến giả cho loại xe

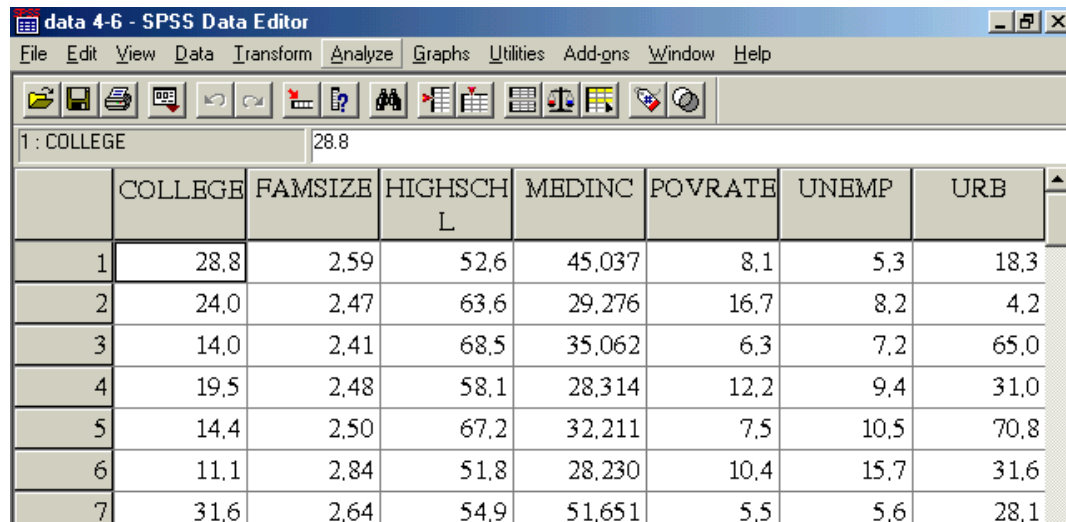
Vào **Transform, Recode, Into Different Variables**. Tức là chúng ta sẽ mã hóa lại biến **loai_xe**, và sẽ cho ra một biến mới (nếu chọn **Into Same Variables** thì SPSS sẽ biến đổi rồi thay thế luôn thông tin của biến cũ).

HÌNH 6



Lấy dữ liệu từ file DATA4-6 của Ramanathan.

HÌNH 8



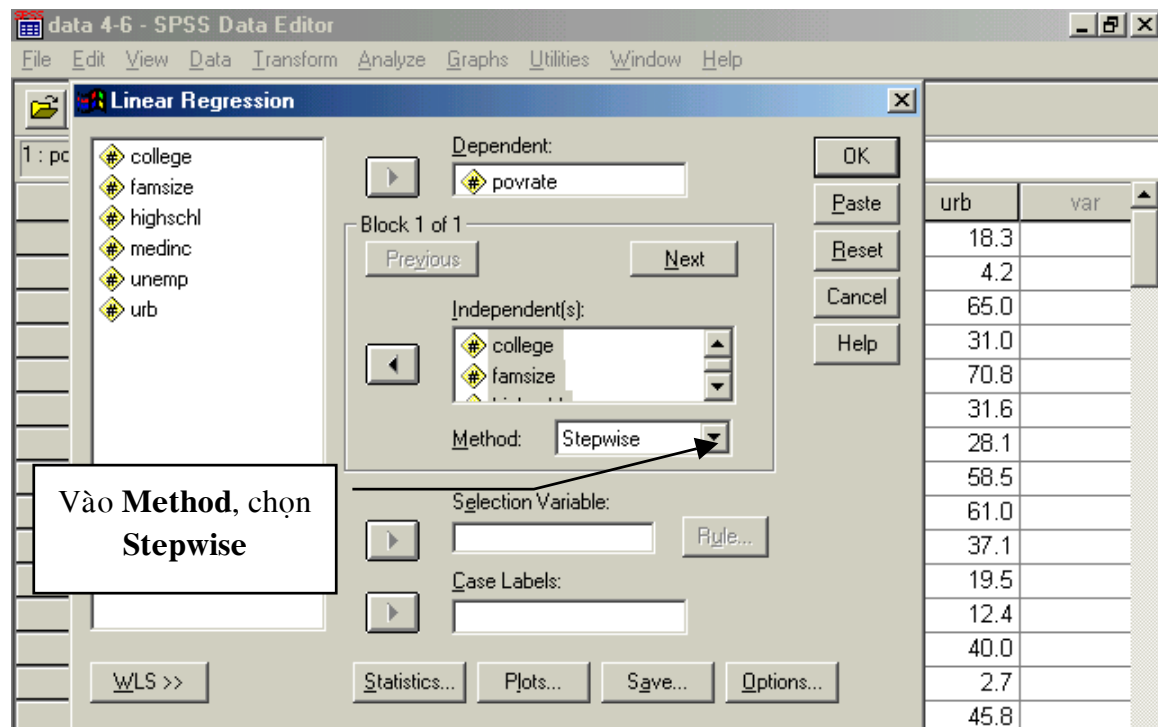
	COLLEGE	FAMSIZE	HIGHSCH L	MEDINC	POVRATE	UNEMP	URB
1	28,8	2,59	52,6	45,037	8,1	5,3	18,3
2	24,0	2,47	63,6	29,276	16,7	8,2	4,2
3	14,0	2,41	68,5	35,062	6,3	7,2	65,0
4	19,5	2,48	58,1	28,314	12,2	9,4	31,0
5	14,4	2,50	67,2	32,211	7,5	10,5	70,8
6	11,1	2,84	51,8	28,230	10,4	15,7	31,6
7	31,6	2,64	54,9	51,651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