

PHÂN TÍCH DỮ LIỆU BẰNG PHẦN MỀM SPSS 12.0*

PHẦN 2

Các nội dung chính trong phần này:

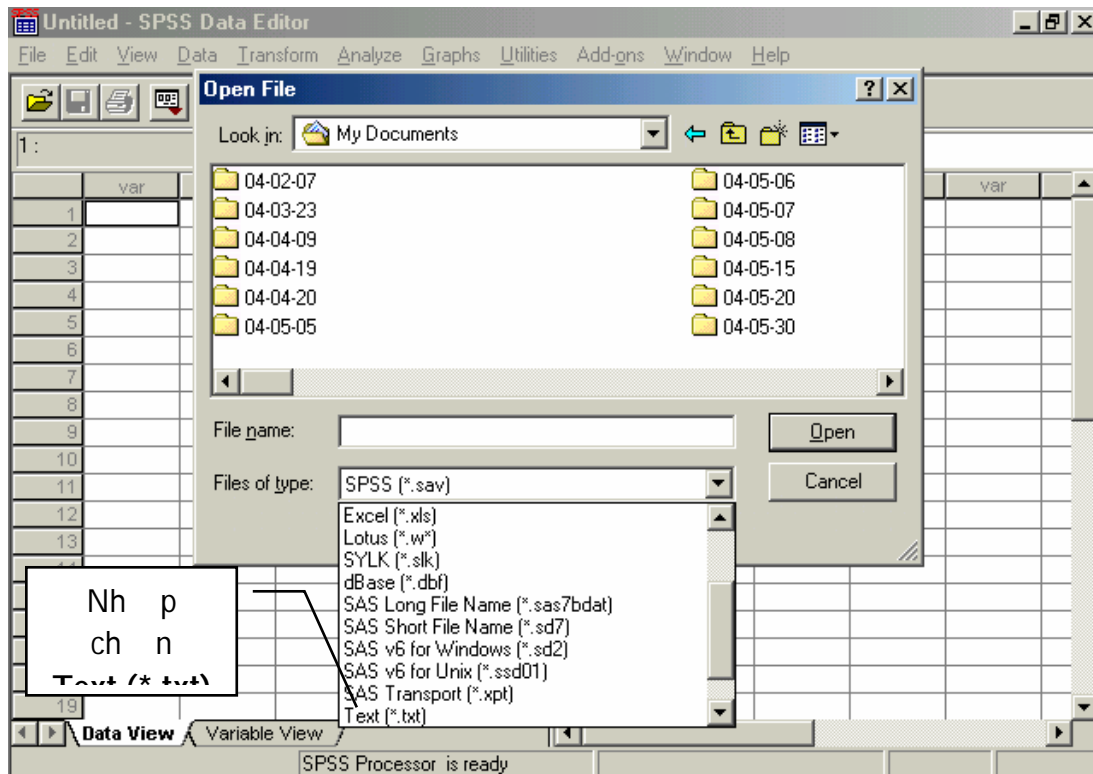
1. Mở dữ liệu từ tập tin dạng text
2. Xem xét sự tương quan giữa các biến
3. Hồi quy OLS (trường hợp đơn biến)
4. Hồi quy OLS (trường hợp đa biến)
5. Hồi quy trong trường hợp có hiện tượng đa cộng tuyến hoàn hảo

* 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).

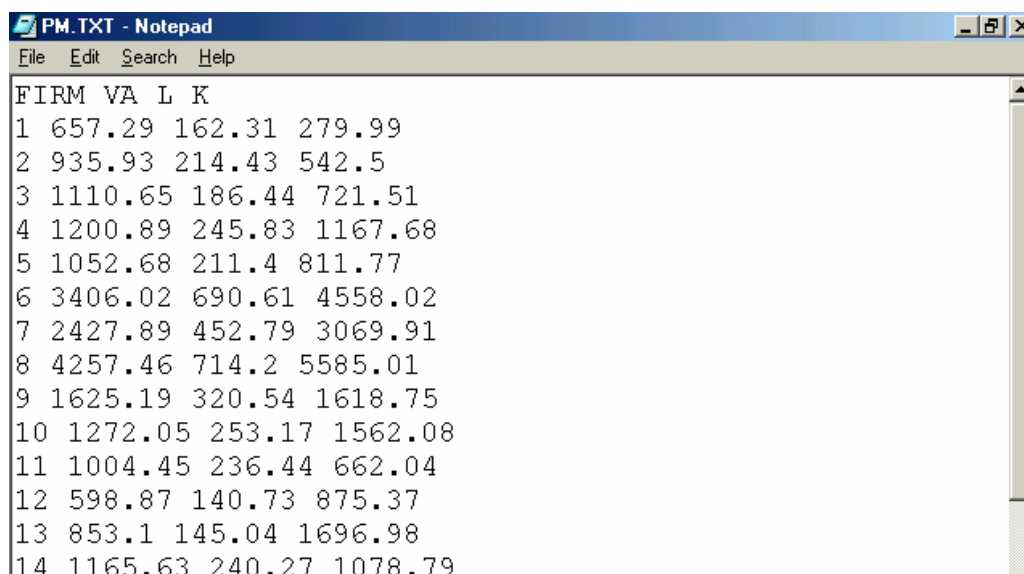
HÌNH 1



Ví dụ:

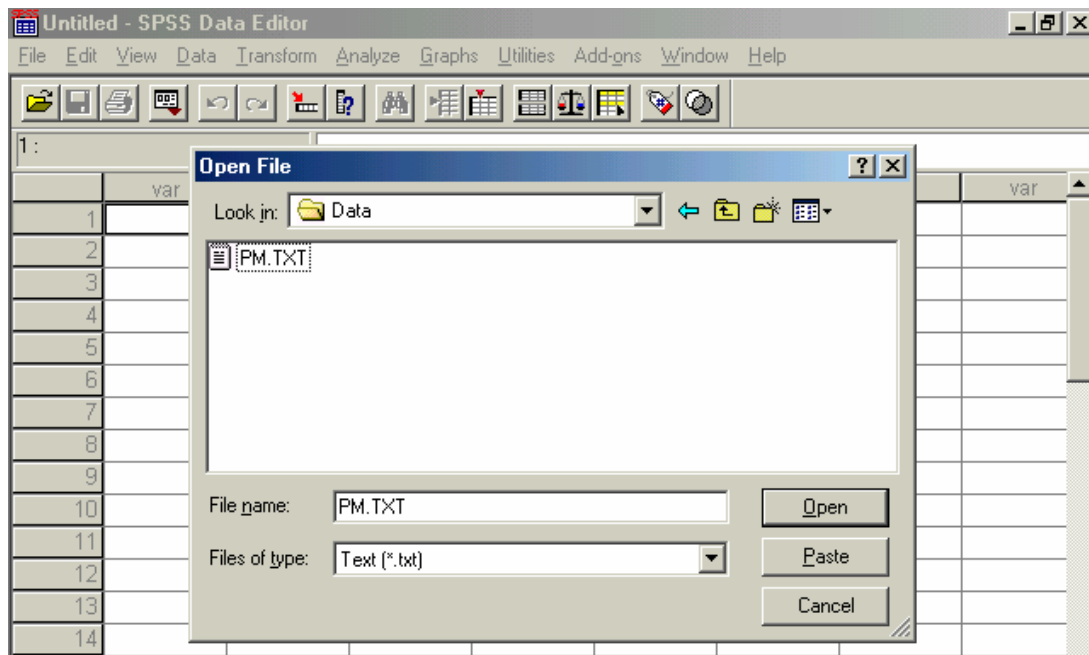
Chúng ta sẽ mở tập tin **pm.txt** chứa dữ liệu về **FIRM**, **VA**, **K** và **L** đã thực hành trong **EVIEWS**.

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**

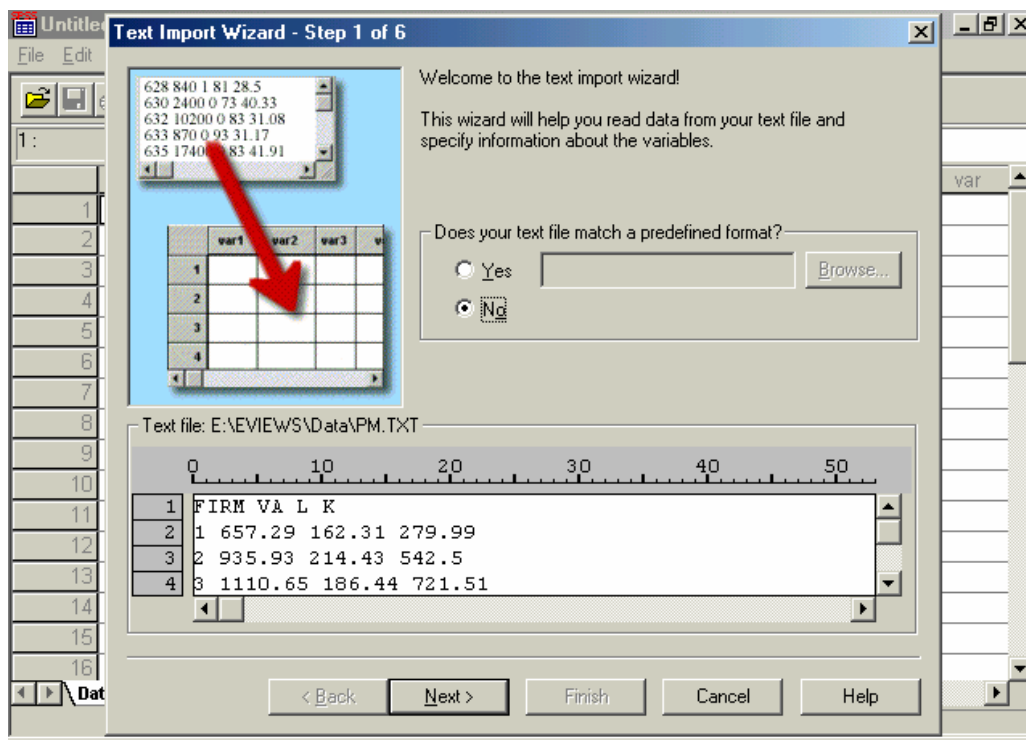
HÌNH 3



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

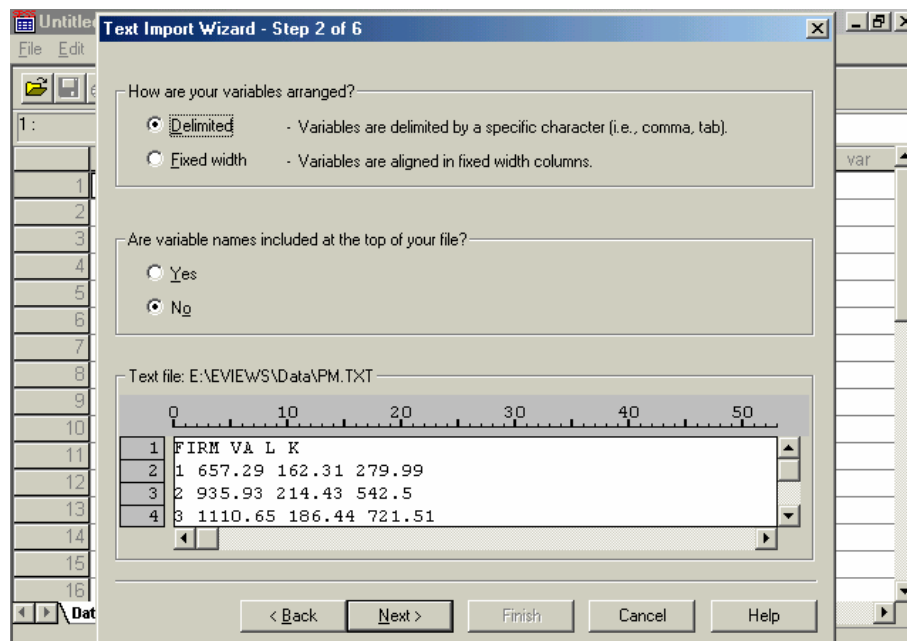
- **Bước 1 (Hình 4):** Trong Bước 1, SPSS mặc định sẵn ở chế độ **No**, chúng ta chỉ việc nhấp **Next**.

HÌNH 4



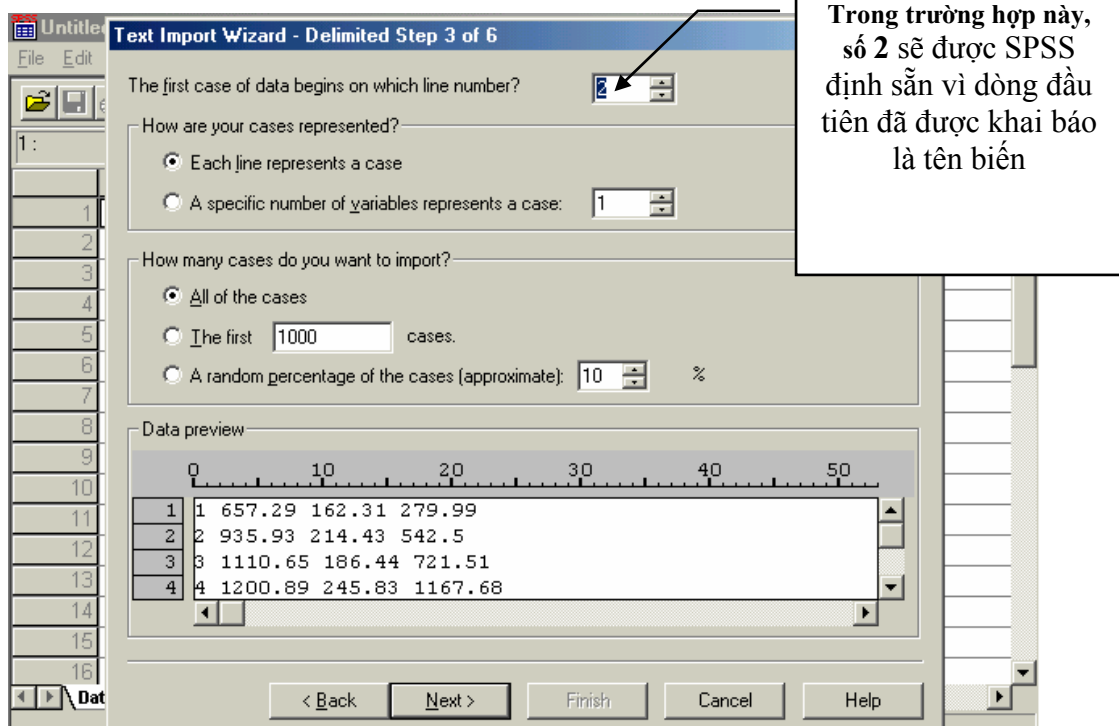
- **Bước 2 (Hình 5):** Trong Bước 2, cần chú ý dòng chữ **Are variable names included at the top of your file?**, ngụ ý hỏi có phải tên biến nằm ở dòng đầu tiên của tập tin đó không? Nếu phải thì nhấp **Yes**, ngược lại thì **No**. Trong trường hợp của tập tin **pm.txt** thì dòng đầu tiên có chứa tên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

HÌNH 5



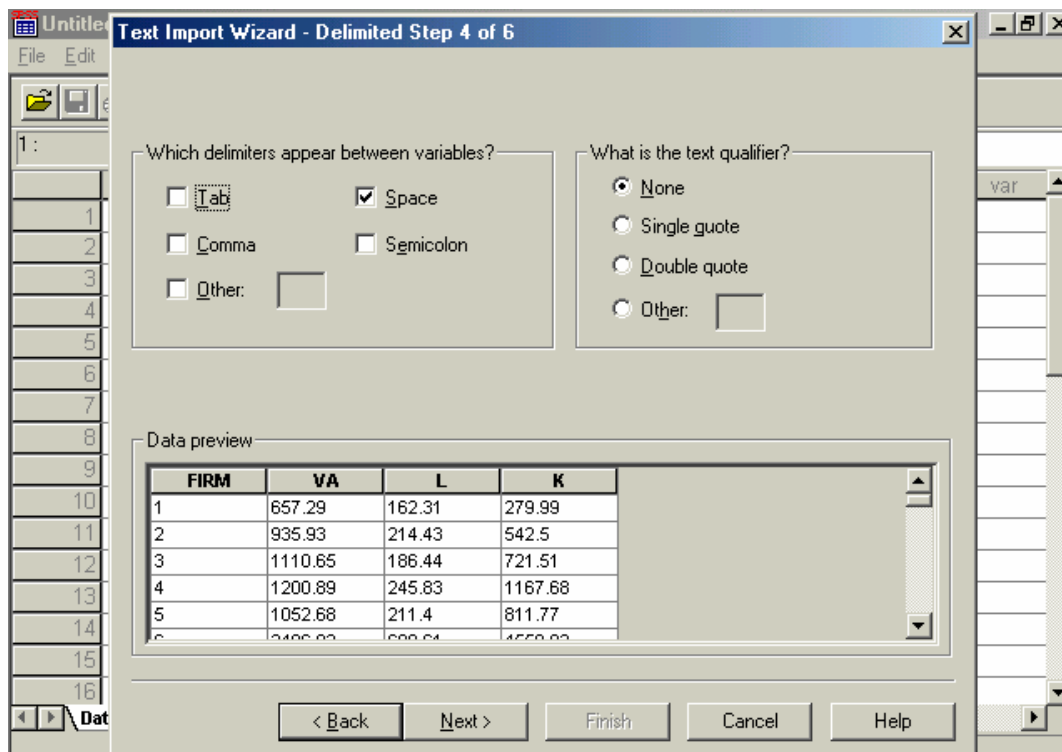
- **Bước 3 (Hình 6):** Chú ý dòng chữ **The first case of data begins on which line number?**, ngụ ý hỏi dữ liệu sẽ bắt đầu từ dòng thứ mấy. Khai báo xong, bấm **Next**.

HÌNH 6



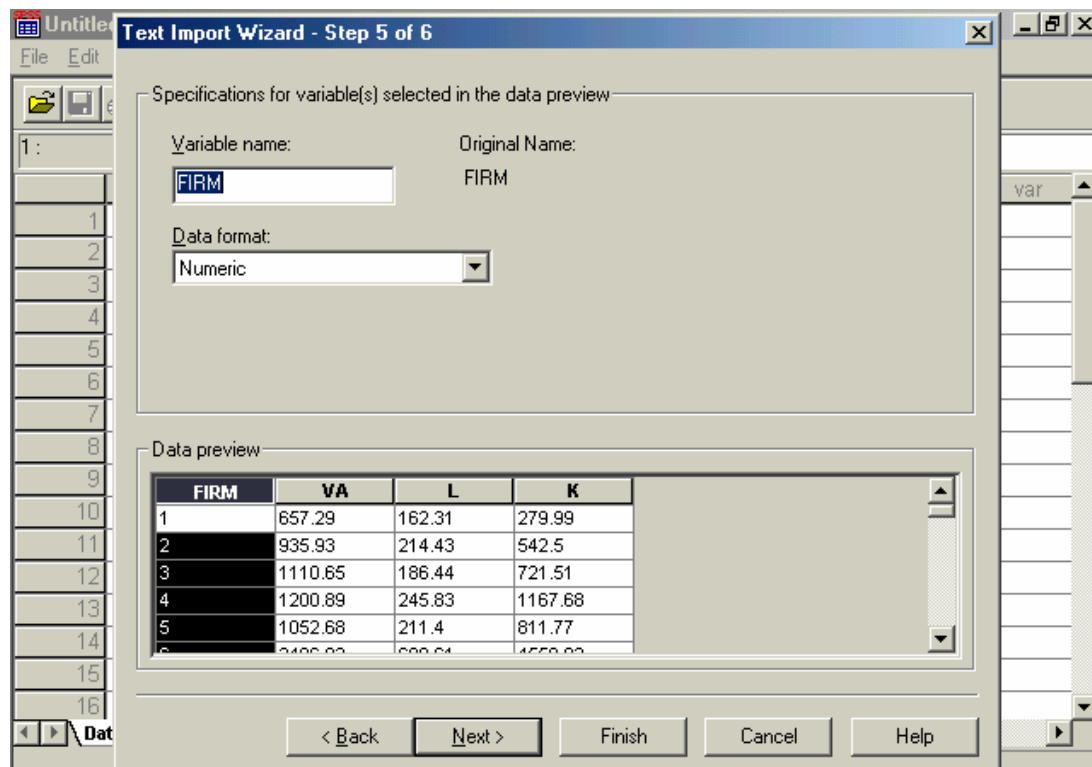
- **Bước 4 (Hình 7):** Chú ý đến câu **Which delimiters appear between variables?**, ngụ ý hỏi dữ liệu của các biến được phân định bằng cách nào. Bằng **Tab**, bằng khoảng trống (**Space**), bằng dấu phẩy (**Comma**), dấu chấm phẩy (**Semicolon**) hay ở dạng cụ thể nào khác (**Other**). Ở đây SPSS thường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

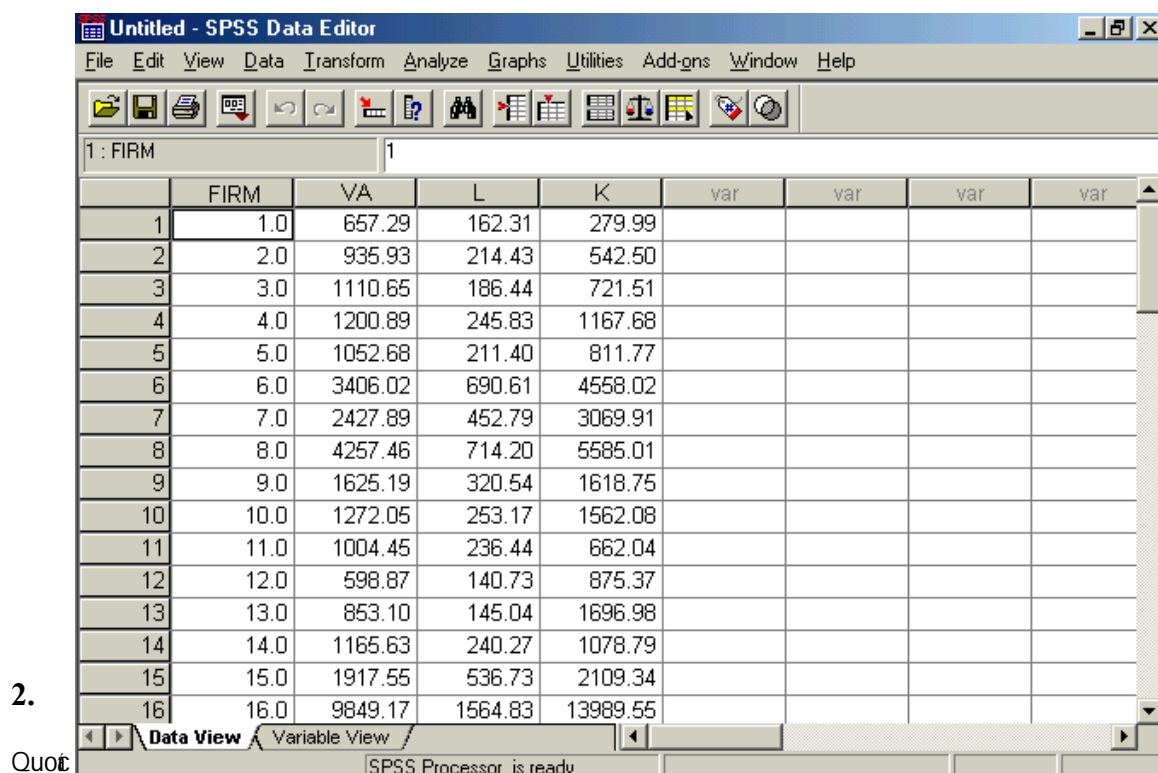
HÌNH 8



- **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



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

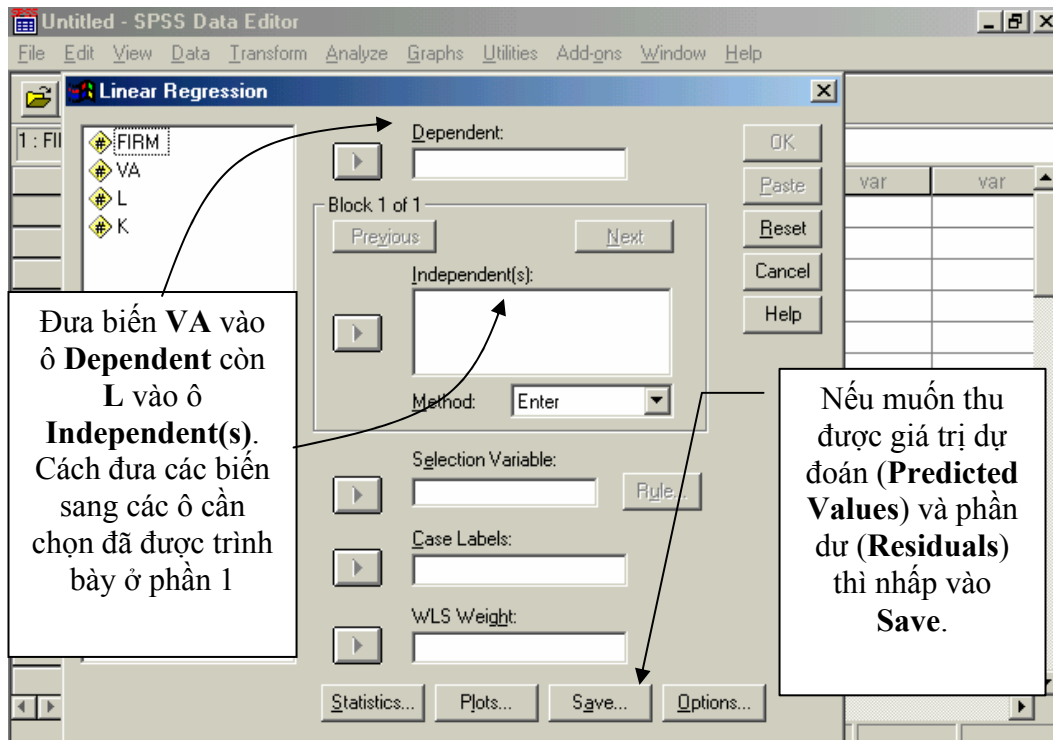
		VA	L	K
VA	Pearson Correlation	1	.965**	.975**
	Sig. (2-tailed)	.	.000	.000
	N	27	27	27
L	Pearson Correlation	.965**	1	.964**
	Sig. (2-tailed)	.000	.	.000
	N	27	27	27
K	Pearson Correlation	.975**	.964**	1
	Sig. (2-tailed)	.000	.000	.
	N	27	27	27

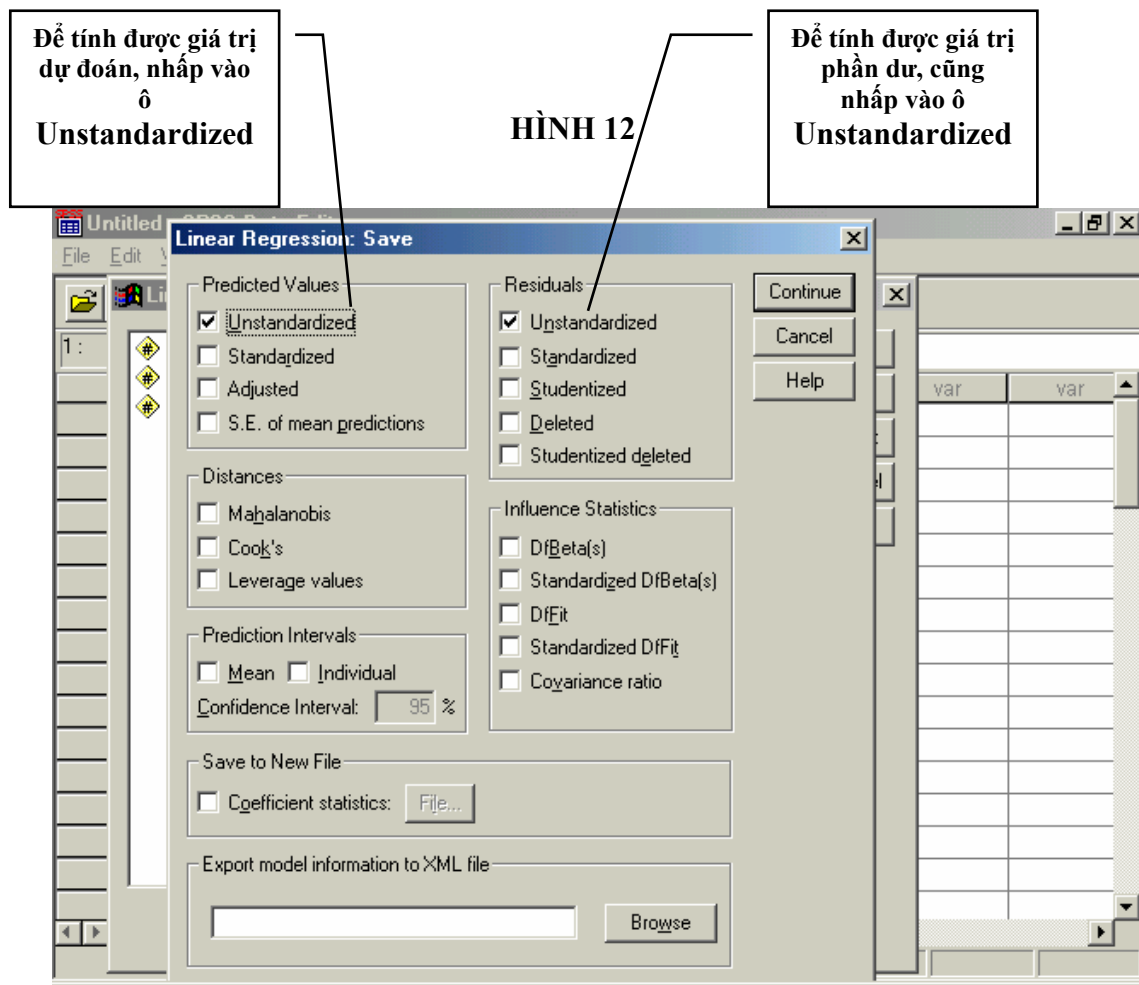
** . Correlation is significant at the 0.01 level (2-tailed).

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





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

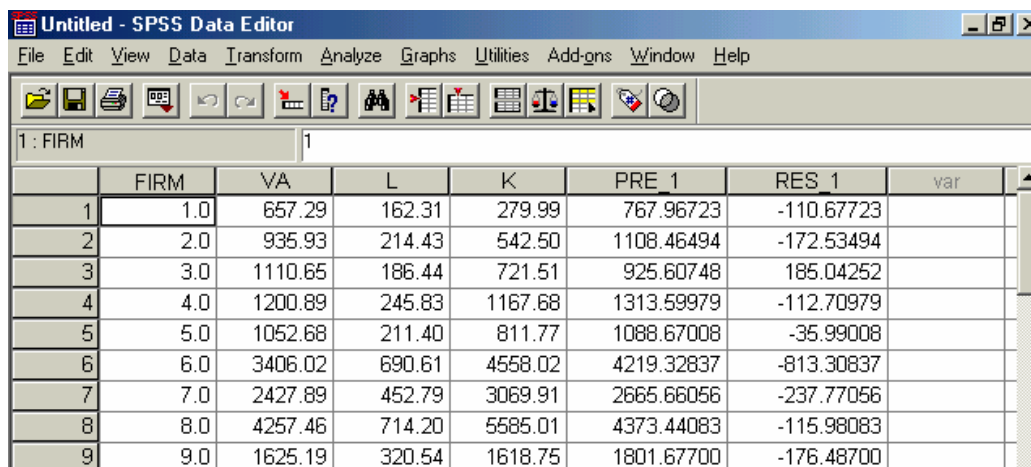
Coefficients^a

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

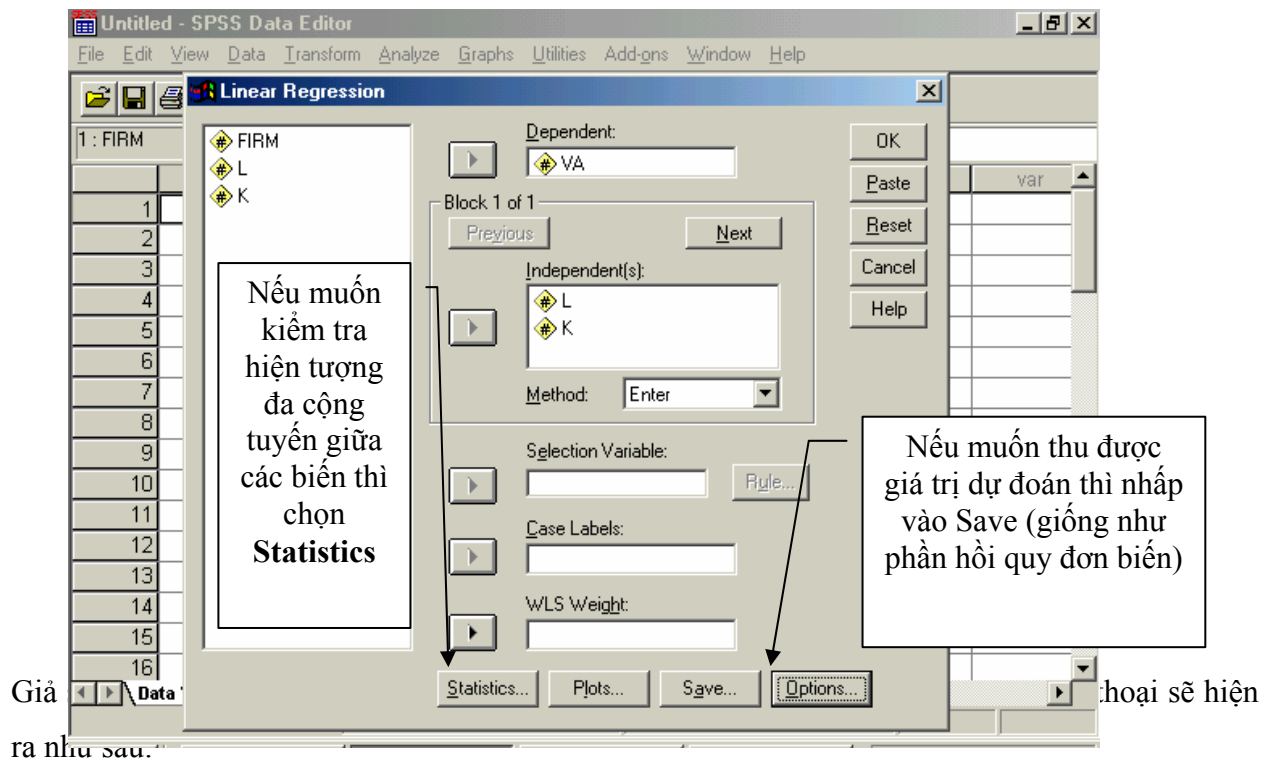


	FIRM	VA	L	K	PRE_1	RES_1	var
1	1.0	657.29	162.31	279.99	767.96723	-110.67723	
2	2.0	935.93	214.43	542.50	1108.46494	-172.53494	
3	3.0	1110.65	186.44	721.51	925.60748	185.04252	
4	4.0	1200.89	245.83	1167.68	1313.59979	-112.70979	
5	5.0	1052.68	211.40	811.77	1088.67008	-35.99008	
6	6.0	3406.02	690.61	4558.02	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	

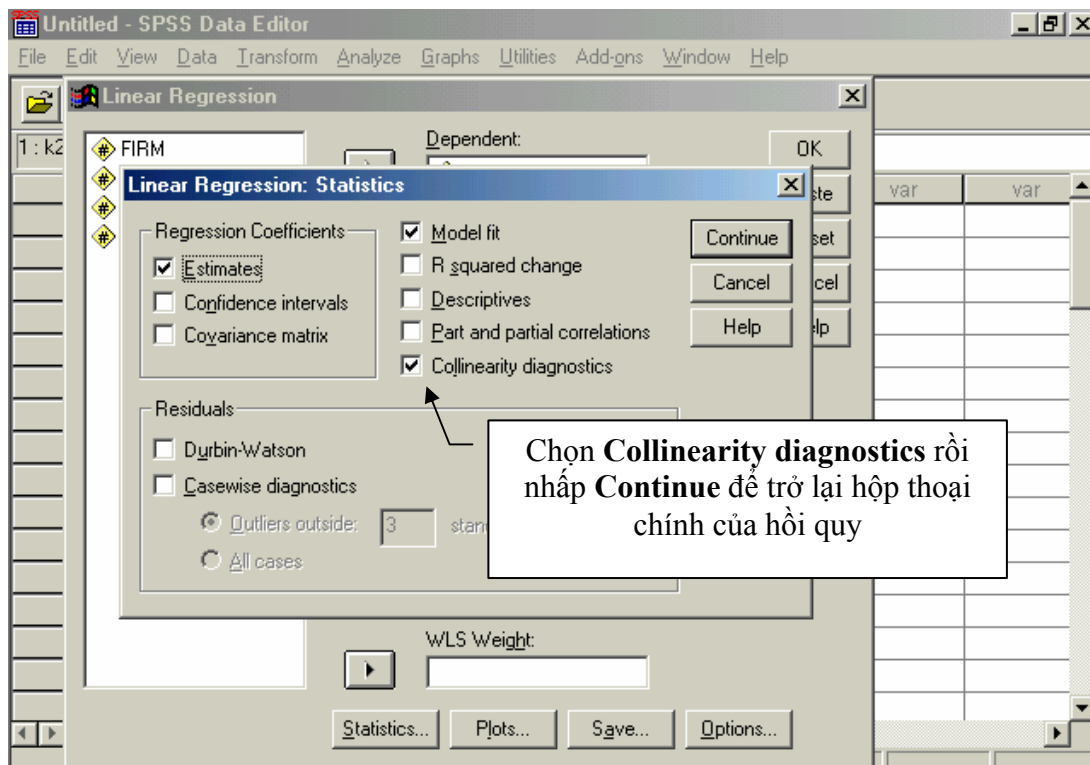
4. Hồi quy OLS (trường hợp đa biến)

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^a

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	. ^a000	.	.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 $K2 = 2 \times K$).