

XỬ LÝ VÀ TRÌNH BÀY SỐ LIỆU

ĐẶNG THỊ THU HƯƠNG – BM CNSTH- K CNTP

THỦ TỤC THỐNG KÊ CƠ BẢN

- Kết quả thí nghiệm cần phải được xử lý thống kê
- **Với thí nghiệm thăm dò cổ điển** (phân tích sự biến động) - kết quả thí nghiệm được thể hiện qua giá trị trung bình và độ lệch chuẩn. Kết quả trình bày dưới dạng bảng hoặc biểu đồ - Cần chỉ rõ các giá trị trung bình có khác biệt (thường $p < 0,05$) thống kê hay không?
- **Với thí nghiệm tối ưu hóa** (thăm dò và khai thác mặt đáp ứng) – kết quả thí nghiệm được thể hiện qua giá trị trung bình và độ lệch chuẩn – Kết quả trình bày dưới dạng phương trình (hàm số), mặt đáp ứng và đường đồng mức. – Cần chỉ rõ các yếu tố có ảnh hưởng đến hàm mục tiêu hay không, mặt đáp ứng đã thăm dò có phù hợp hay không, Nếu câu trả lời là có thì tiến hành bước tối ưu hóa như thế nào?

THỦ TỤC THỐNG KÊ CƠ BẢN

- Nếu số liệu có phân phối chuẩn (thường số liệu thu được trong phòng thí nghiệm có phân phối chuẩn):

- Khi cần so sánh 2 giá trị trung bình, áp dụng t-test
- Khi cần so sánh từ 3 giá trị trung bình trở lên, áp dụng ANOVA và Post hoc test (phân tích sâu ANOVA)

Nếu số liệu không có phân phối chuẩn (thường số liệu điều tra, khảo sát, thu được từ thực địa không có phân phối chuẩn):

- Sử dụng các kiểm định phi tham số

LỰA CHỌN PHẦN MỀM XỬ LÝ SỐ LIỆU

- Người nghiên cứu cần có kiến thức thống kê cơ bản, biết cách thiết kế thí nghiệm hợp lý và lựa chọn thủ tục thống kê phù hợp để xử lý số liệu.
- Biết cách sử dụng các phần mềm xử lý thống kê để hỗ trợ

Tiêu chí lựa chọn phần mềm:

Giao diện thân thiện, dễ sử dụng, nhanh, đơn giản (Excel, SPSS, SAS, STATITICA, STATGRAPHICS, MINITAB, R, SIGMASTAT, DX,...) – Có những phần mềm miễn phí, phần mềm bản quyền (bản dùng thử, bản crack) – Tất cả các phần mềm đều xử lý được số liệu với những thủ tục thống kê cơ bản, và có điểm mạnh về một mảng nào đó.

Người nghiên cứu tự chọn cho mình một phần mềm phù hợp và khi trình bày báo cáo, bài báo khoa học cần chỉ rõ: số liệu đã được xử lý và biểu đồ, đồ thị đã được vẽ bằng phần mềm nào (**nêu nguồn gốc (tác giả bản quyền của phần mềm đó), phiên bản, thủ tục thống kê nào đã được áp dụng,..**)

THU THẬP SỐ LIỆU

- Khi thu thập số liệu, người nghiên cứu cần xác định trước phần mềm xử lý số liệu mà mình sẽ sử dụng, thủ tục thống kê mà mình sẽ áp dụng để có cách nhập số liệu cho hợp lý.
- Các số liệu nên nhập trước trong file Excel (thường nhập theo cột) – Xem file excel đính kèm

Theo mục tiêu của tình huống nghiên cứu, tác giả chọn MS-Excel 2013, STATISTICA để vẽ biểu đồ và xử lý số liệu. Thủ tục thống kê cần áp dụng:

- Tính giá trị trung bình và độ lệch chuẩn (thống kê mô tả (Descriptive Statistic trong Excel hoặc STATISTICA hay đơn giản là dùng hàm AVERAGE và STDEV trong excel)).
- So sánh các giá trị trung bình để chỉ ra sự khác biệt có ý nghĩa thống kê: Do có 3 nhóm mẫu (A, B, C) nên chọn ONE WAY ANOVA – VÀ TEST POST HOC để kiểm tra sự khác biệt giữa các giá trị trung bình
- Để đánh giá mối tương quan giữa chỉ tiêu đánh giá và thời gian bảo quản đồng thời biết được xu hướng tăng (giảm) của chỉ tiêu đánh giá theo thời gian – sử dụng Linear Regression

Bảng số liệu nhập trong excel để xử lý thống kê

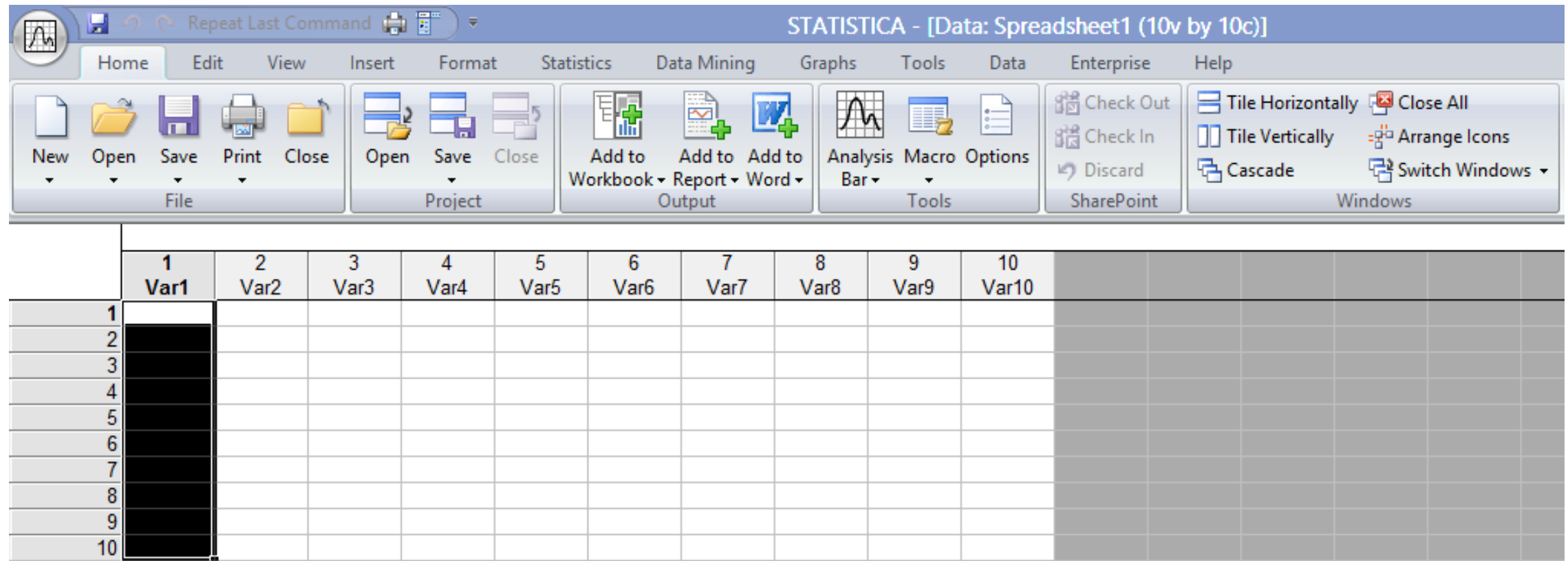
A	B	C	D	E	F	G	H	I	J
Mẫu	Y0	Y1	Y2	Y3	Y4	THOI GIAN	YA	YB	YC
A	47.22	49.19	51.61	53.31	55.55	0	47.22	49.00	49.53
A	48.63	49.71	51.73	53.60	55.24	0	48.63	49.41	48.74
A	46.87	49.27	51.82	53.30	55.97	0	46.87	49.01	49.00
B	49.00	44.03	43.25	40.24	36.63	1	49.19	44.03	47.80
B	49.41	44.60	43.56	40.73	36.25	1	49.71	44.60	47.95
B	49.01	44.95	43.37	40.37	36.64	1	49.27	44.95	47.79
C	49.53	47.80	45.96	40.40	39.72	2	51.61	43.25	45.96
C	48.74	47.95	45.74	40.82	39.39	2	51.73	43.56	45.74
C	49.00	47.79	45.87	40.87	39.68	2	51.82	43.37	45.87
						3	53.31	40.24	40.40
						3	53.60	40.73	40.82
Y0, 1,2,3,4: Giá trị của chỉ tiêu đánh giá ở tuần 0, 1,2,3,4						3	53.30	40.37	40.87
YA, YB, YC: Giá trị của chỉ tiêu đánh giá của các mẫu A,B,C						4	55.55	36.63	39.72
						4	55.24	36.25	39.39
						4	55.97	36.64	39.68

HƯỚNG DẪN SỬ DỤNG STATISTICA

**Tác giả sử dụng STATISTICA software (Version
10.0, StatSoft, Inc. 2300 East 14th Street Tulsa,
OK 74104 USA);**

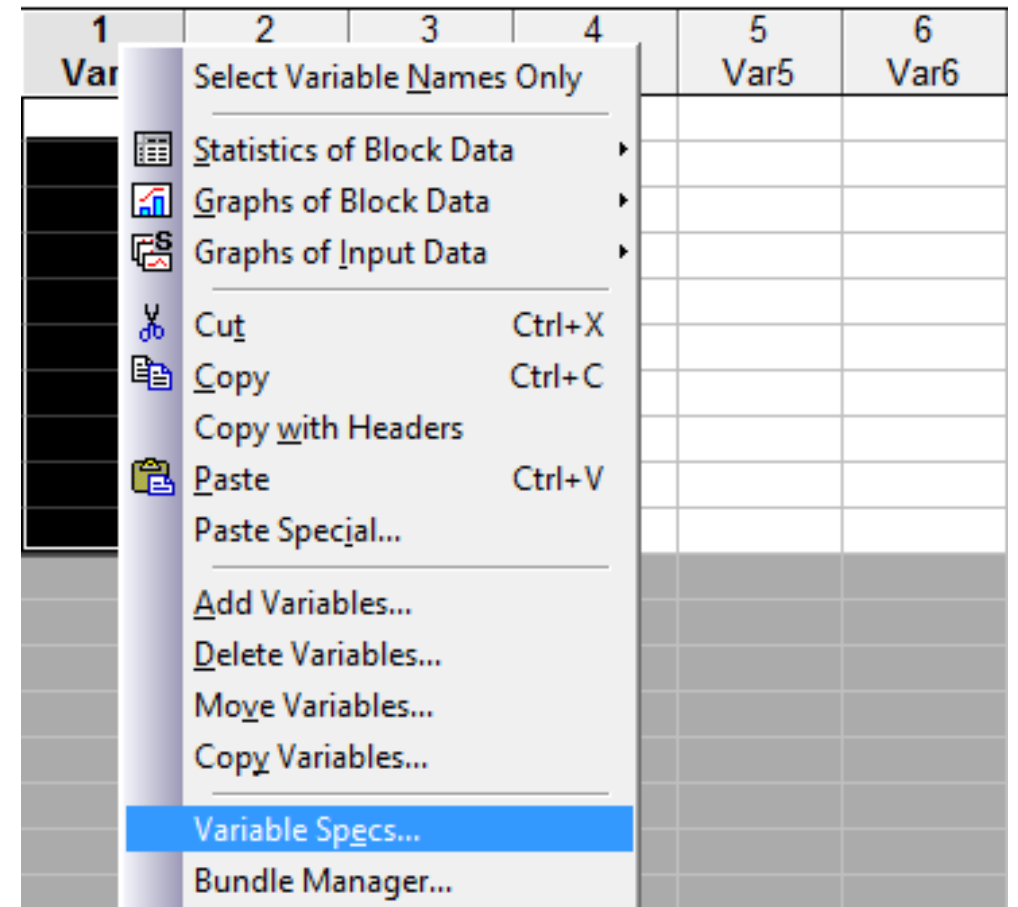
Tình huống áp dụng (file EXCEL đính kèm)

GIAO DIỆN CỦA STATISTICA



BƯỚC 1: KHAI BÁO BIẾN

Nhấp chuột phải vào vị trí biến cần khai báo - nếu khai báo mới thường bắt đầu từ Var1- chọn Variable Spec...



BƯỚC 2

Tại cửa sổ khai báo thông tin biến

- Nhấp chuột vào vị trí All Specs... để khai báo nhiều biến cùng một lúc- Tên biến có thể chọn general hoặc number
- Chỉ tiêu đánh giá chọn dữ liệu number

The screenshot shows the 'Variable 1' dialog box in SPSS. The 'Name' field contains 'Var1'. The 'Type' is set to 'Double'. The 'Measurement Type' is 'Auto'. The 'Length' is '8'. The 'MD code' is '-999999998'. The 'Display format' list is open, showing options: General, Number, Date, Time, Scientific, Currency, Percentage, Fraction, and Custom. The 'All Specs...' button is highlighted. The 'Long name (label or formula with Functions):' field is empty, and the 'Function guide' checkbox is checked.

BƯỚC 3
Khai báo cùng lúc tất cả các biến

1	2	3	4	5	6	7	8	9	10
MAU	Y0	Y1	Y2	Y3	Y4	THOI GIAN	YA	YB	YC

Variable Specifications Editor									
<div> A <div>Arial</div> <div>10</div> <div> B <i>I</i> <u>U</u> </div> <div> A ✂ 📄 📋 </div> <div>Vars</div> </div>									
	Name	Type	MD code	Length	Long Name (label or formula)	Measurement Type	Ex		
1	MAU	Double	-999999998			Auto			
2	Y0	Double	-999999998			Auto			
3	Y1	Double	-999999998			Auto			
4	Y2	Double	-999999998			Auto			
5	Y3	Double	-999999998			Auto			
6	Y4	Double	-999999998			Auto			
7	THOI GIAN	Double	-999999998			Auto			
8	YA	Double	-999999998			Auto			
9	YB	Double	-999999998			Auto			
10	YC	Double	-999999998			Auto			

Nhập số liệu vào STATISTICA

STATISTICA - [Data: SO LIEU THO 2552014]

Home Edit View Insert Format Statistics Data Mining Graphs Tools Data Enterprise

Basic Statistics Regression ANOVA Nonparametrics Distribution Fitting More Distributions

Advanced Models Mult/Exploratory Power Analysis Neural Nets PLS, PCA, ... Variance

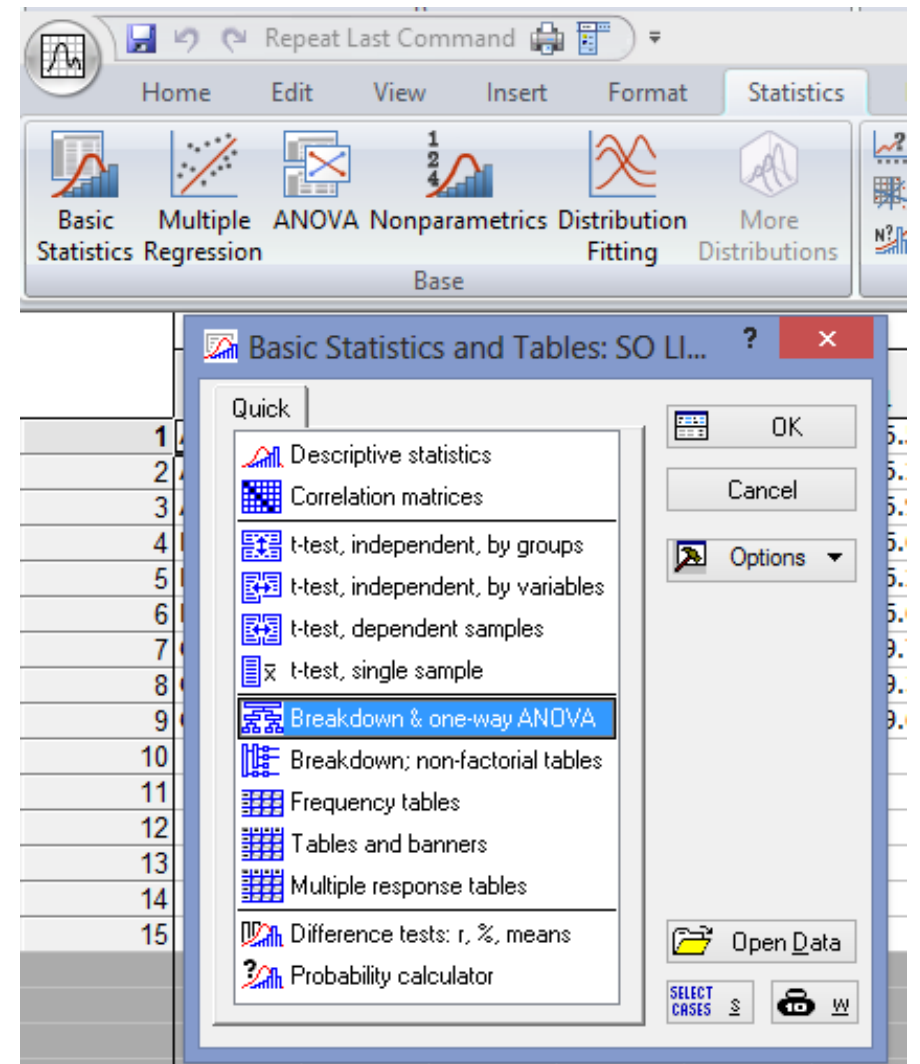
QC Charts Multivariate Predictive

	1 MAU	2 Y0	3 Y1	4 Y2	5 Y3	6 Y4	7 THOI GIAN	8 YA	9 YB	10 YC	11
1	A	47.22	49.19	51.61	53.31	55.55	0	47.22	49.00	49.53	
2	A	48.63	49.71	51.73	53.60	55.24	0	48.63	49.41	48.74	
3	A	46.87	49.27	51.82	53.30	55.97	0	46.87	49.01	49.00	
4	B	49.00	44.03	43.25	40.24	36.63	1	49.19	44.03	47.80	
5	B	49.41	44.60	43.56	40.73	36.25	1	49.71	44.60	47.95	
6	B	49.01	44.95	43.37	40.37	36.64	1	49.27	44.95	47.79	
7	C	49.53	47.80	45.96	40.40	39.72	2	51.61	43.25	45.96	
8	C	48.74	47.95	45.74	40.82	39.39	2	51.73	43.56	45.74	
9	C	49.00	47.79	45.87	40.87	39.68	2	51.82	43.37	45.87	
10							3	53.31	40.24	40.40	
11							3	53.60	40.73	40.82	
12							3	53.30	40.37	40.87	
13							4	55.55	36.63	39.72	
14							4	55.24	36.25	39.39	
15							4	55.97	36.64	39.68	

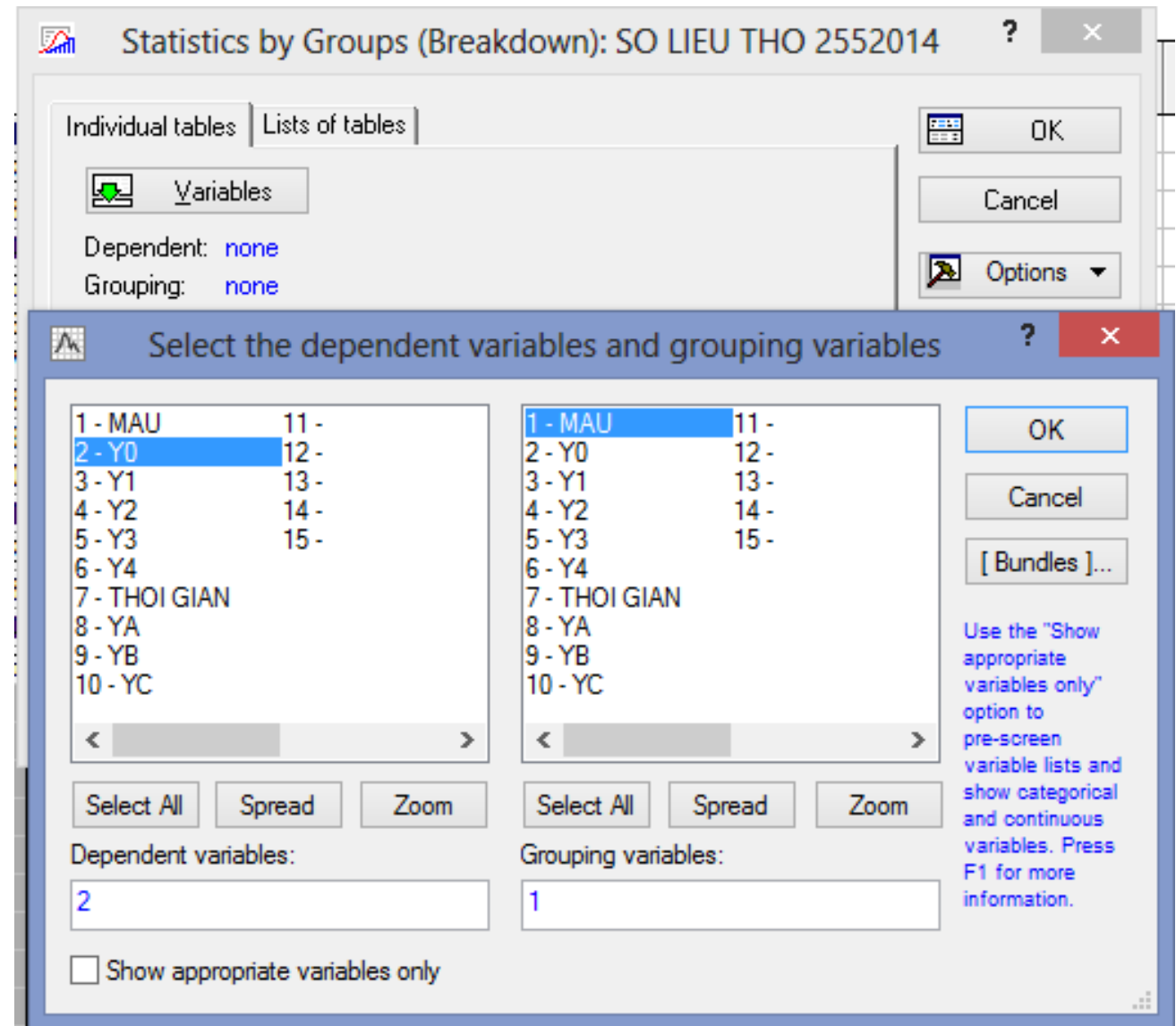
- Về mặt lý thuyết, cần phân tích one way ANOVA để biết có sự khác biệt có ý nghĩa giữa các giá trị trung bình hay không, nếu có sẽ tiếp tục phân tích Post Hoc để chỉ cụ thể sự khác biệt.
- Với STATISTICA - Sự khác biệt giữa các giá trị trung bình được thể hiện trong bảng kết quả (màu đỏ)
- Khi phân tích phương sai một yếu tố:
- Các nhóm so sánh phải độc lập và được chọn một cách ngẫu nhiên.
- Các nhóm so sánh phải có phân phối chuẩn hoặc cỡ mẫu phải đủ lớn để được xem như tiệm cận phân phối chuẩn.
- Phương sai của các nhóm so sánh phải đồng nhất

Để biết được có sự khác biệt về chỉ tiêu đánh giá giữa các mẫu trong cùng một thời gian bảo quản hay không, lần lượt từ tuần 0,1,2,3,4 thực hiện như sau:

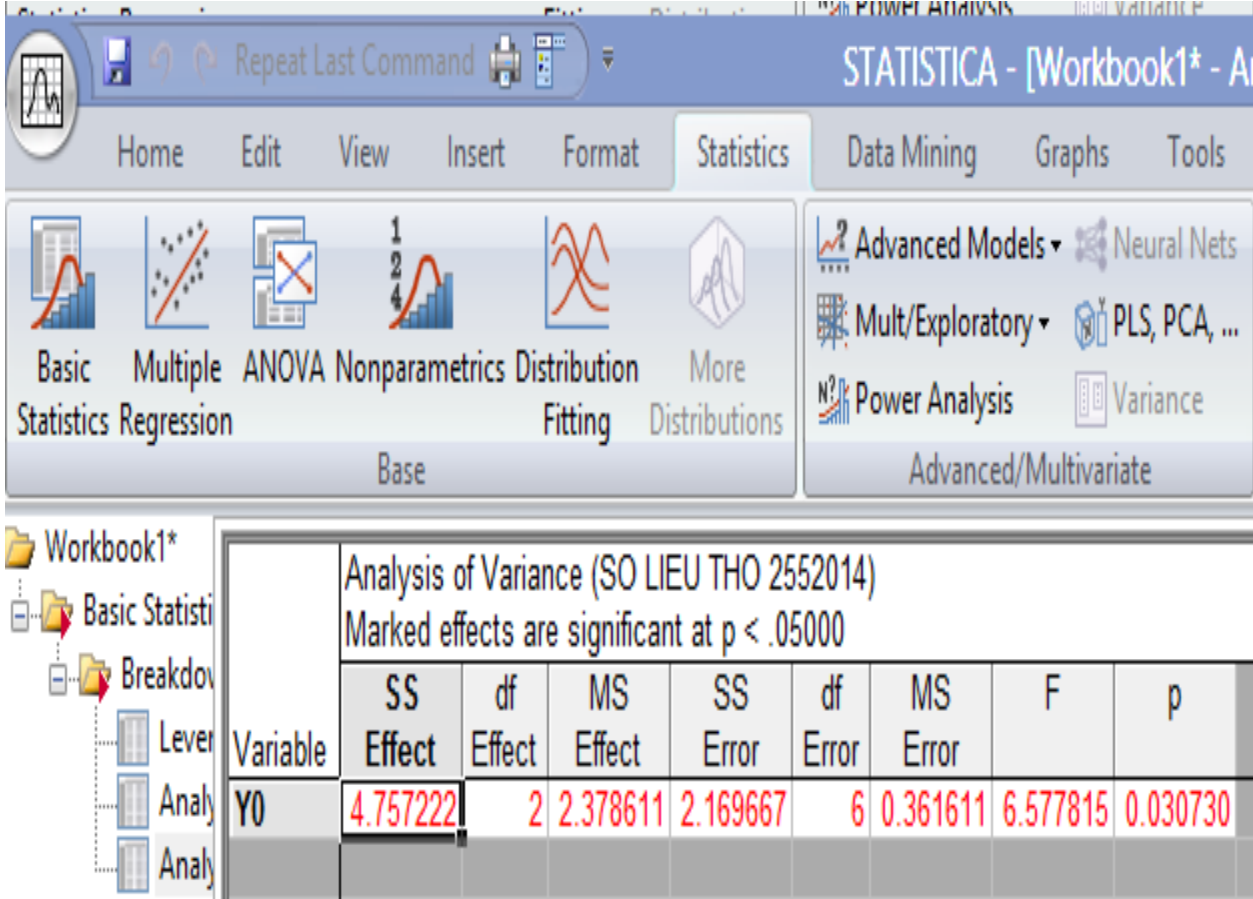
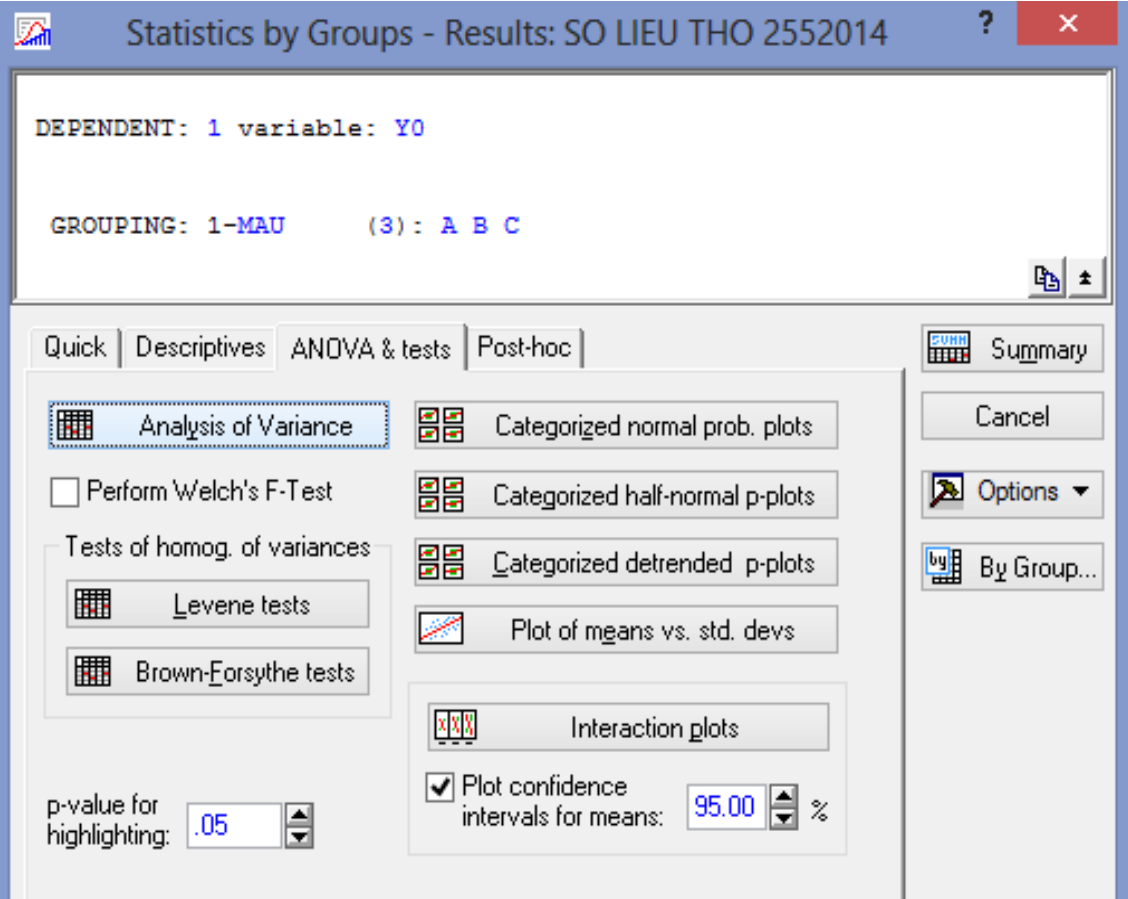
- Chọn statistics- basic statistics- breakdown & one-way ANOVA-OK.



- Trong hộp breakdown – chọn Variables – phần biến phụ thuộc (dependent) chọn Y0, nhóm biến cần so sánh chọn MÃU- OK- OK

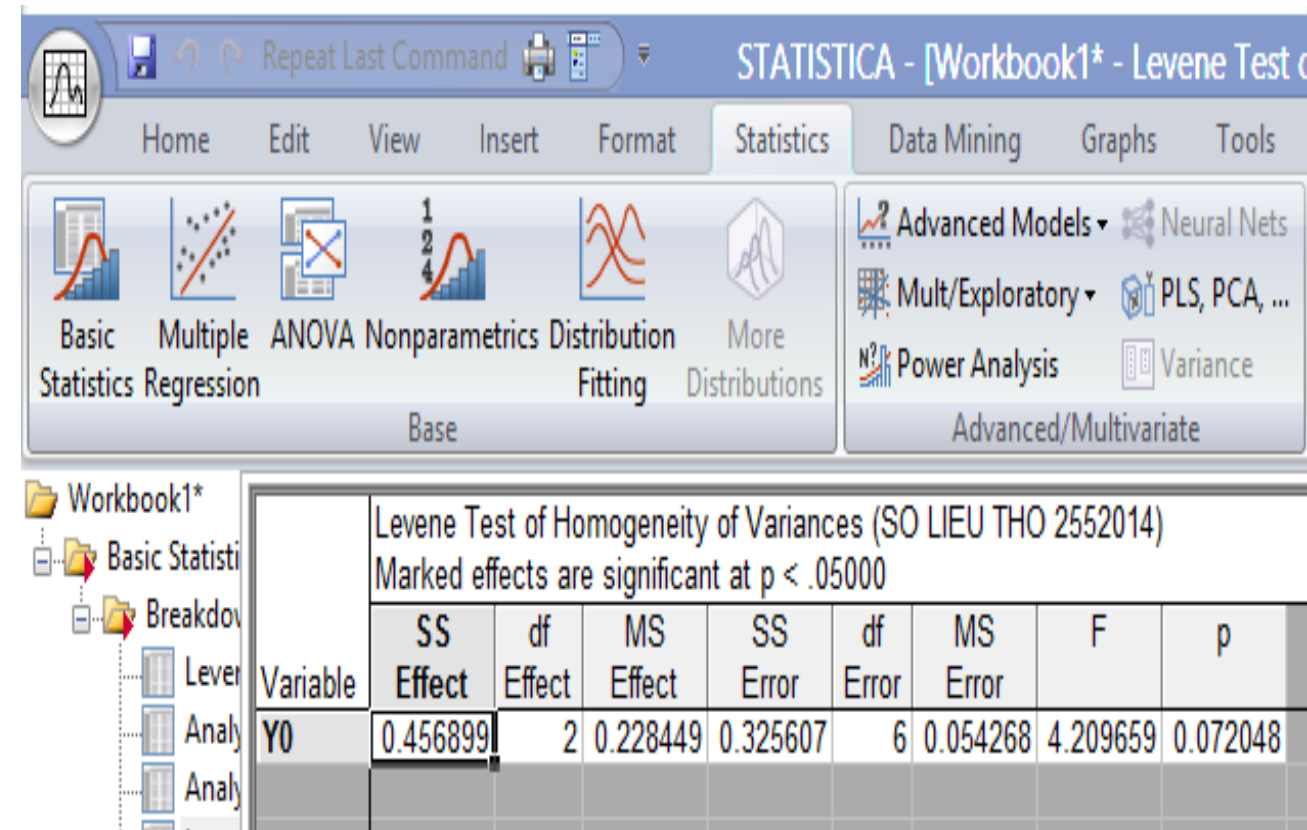
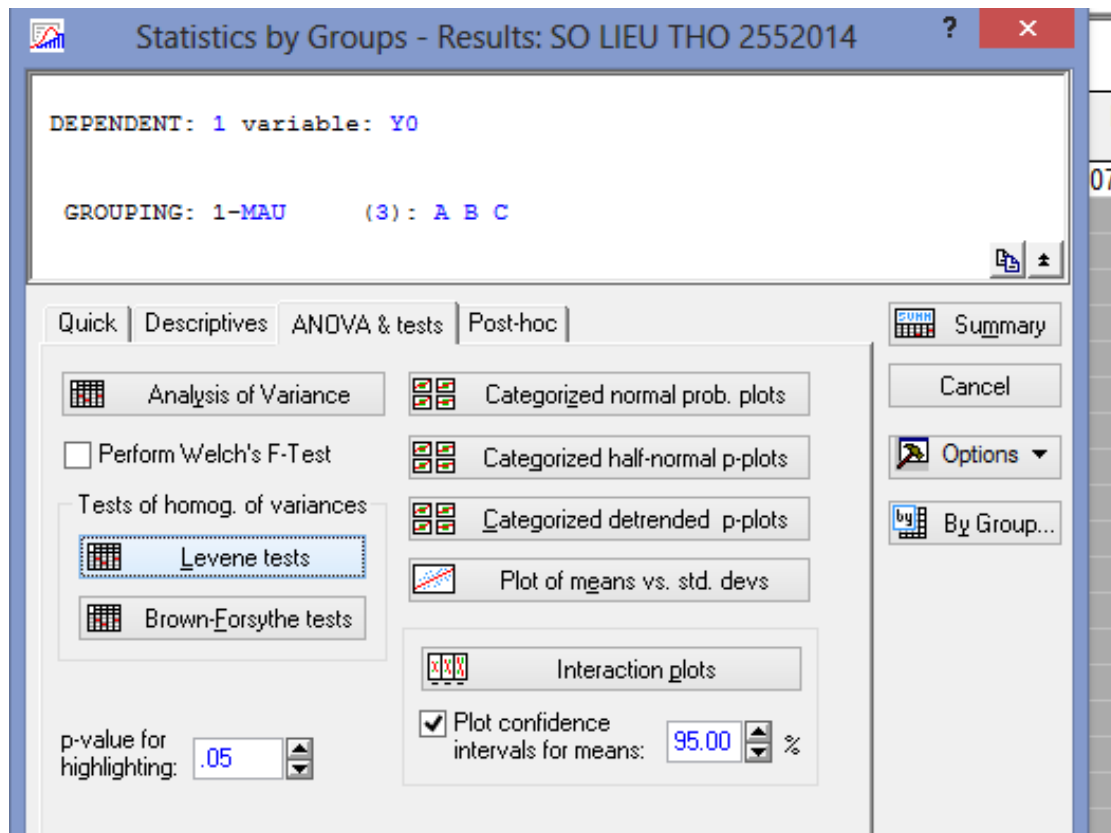


Trong hộp STATISTICS by Groups chọn Analysis of Variance. Kết quả cho thấy có sự khác biệt thống kê giữa các giá trị trung bình (chữ màu đỏ), $p=0,03$

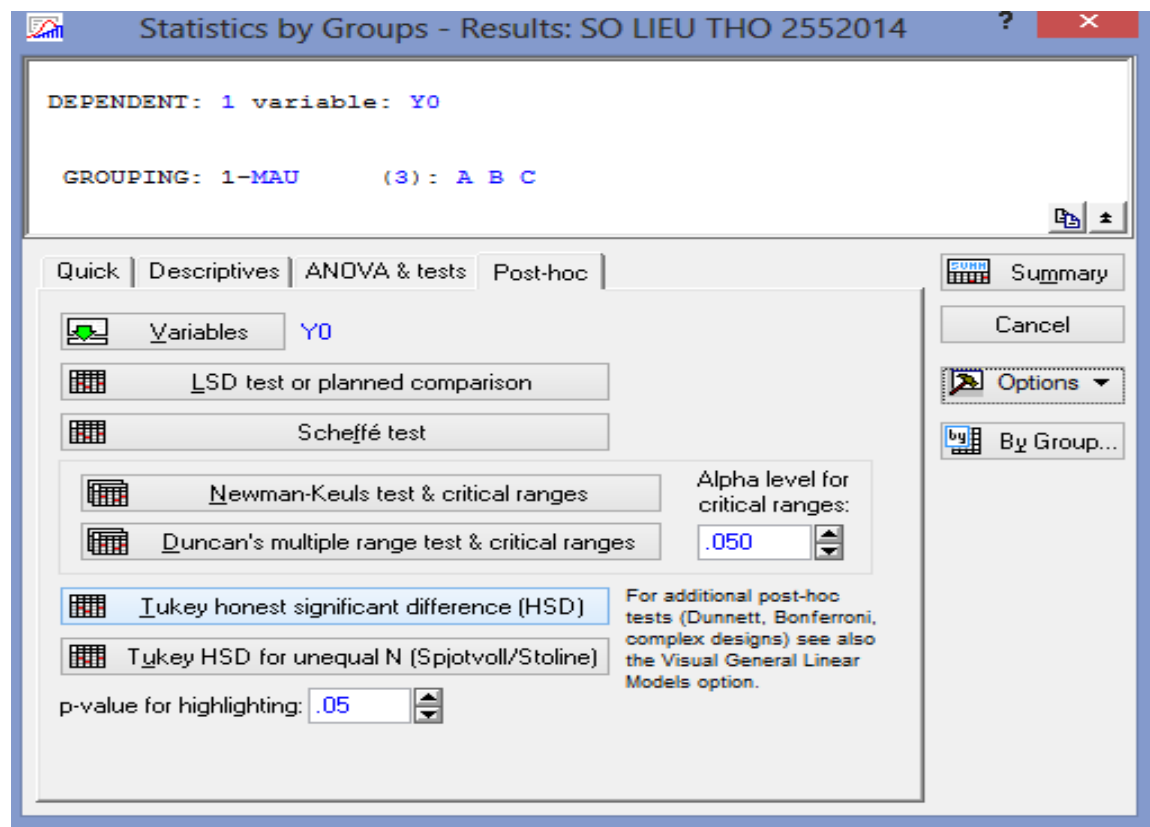


Trong hộp STATISTICS by Groups chọn Levene tests để kiểm tra xem phương sai của 3 nhóm mẫu là đồng nhất hay không?

Kết quả cho thấy có sự đồng nhất giữa các phương sai, $p=0,07$



- Trong hộp Post Hoc test chọn Test TUKEY HD TEST. Sử dụng kiểm định nào là tùy thuộc vào mục đích của nhà nghiên cứu và tình hình thực tế nghiên cứu. Khi trình bày kết quả cần nói rõ kết quả đưa ra được thực hiện bởi kiểm định nào
- Thực hiện tương tự với các mốc thời gian, 1, 2,3,4 tuần,



Tukey HSD test; Variable: Y0 (SO LIEU THO 2552014)
Marked differences are significant at $p < .05000$

		{1}	{2}	{3}
MAU		M=47.573	M=49.140	M=49.090
A	{1}		0.043167	0.048869
B	{2}	0.043167		0.994399
C	{3}	0.048869	0.994399	

Tukey HSD test; Variable: Y0 (Spreadsheet1)					
Marked differences are significant at $p < .05000$					
MAU		{1}	{2}	{3}	
		M=47.573	M=49.140	M=49.090	
A	{1}		0.043167	0.048869	
B	{2}	0.043167		0.994399	
C	{3}	0.048869	0.994399		

Tuần 0: Chỉ tiêu đánh giá của mẫu A khác biệt có ý nghĩa ($p < 0.05$) so với mẫu B và C
 - Mẫu B và C không khác nhau

Tukey HSD test; Variable: Y1 (Spreadsheet1)					
Marked differences are significant at $p < .05000$					
MAU		{1}	{2}	{3}	
		M=49.390	M=44.527	M=47.847	
A	{1}		0.000227	0.002592	
B	{2}	0.000227		0.000240	
C	{3}	0.002592	0.000240		

Tuần 1: Chỉ tiêu đánh giá của 3 mẫu A, B, C hoàn toàn khác biệt ($p < 0.05$)

Tukey HSD test; Variable: Y2 (Spreadsheet1)					
Marked differences are significant at $p < .05000$					
MAU		{1}	{2}	{3}	
		M=51.720	M=43.393	M=45.857	
A	{1}		0.000227	0.000227	
B	{2}	0.000227		0.000227	
C	{3}	0.000227	0.000227		

Tuần 2: Chỉ tiêu đánh giá của 3 mẫu A, B, C hoàn toàn khác biệt ($p < 0.05$)

Tukey HSD test; Variable: Y3 (Spreadsheet1)					
Marked differences are significant at $p < .05000$					
MAU		{1}	{2}	{3}	
		M=53.403	M=40.447	M=40.697	
A	{1}		0.000227	0.000227	
B	{2}	0.000227		0.433257	
C	{3}	0.000227	0.433257		

Tuần 3: Chỉ tiêu đánh giá của mẫu A khác biệt có ý nghĩa ($p < 0.05$) so với mẫu B và C
 - Mẫu B và C không khác nhau

Tukey HSD test; Variable: Y4 (Spreadsheet1)					
Marked differences are significant at $p < .05000$					
MAU		{1}	{2}	{3}	
		M=55.587	M=36.507	M=39.597	
A	{1}		0.000227	0.000227	
B	{2}	0.000227		0.000232	
C	{3}	0.000227	0.000232		

Tuần 4: Chỉ tiêu đánh giá của 3 mẫu A, B, C hoàn toàn khác biệt ($p < 0.05$)

Thực hiện tương tự để xác định xem có sự khác biệt về chỉ tiêu đánh giá trong cùng một mẫu theo thời gian bảo quản hay không.

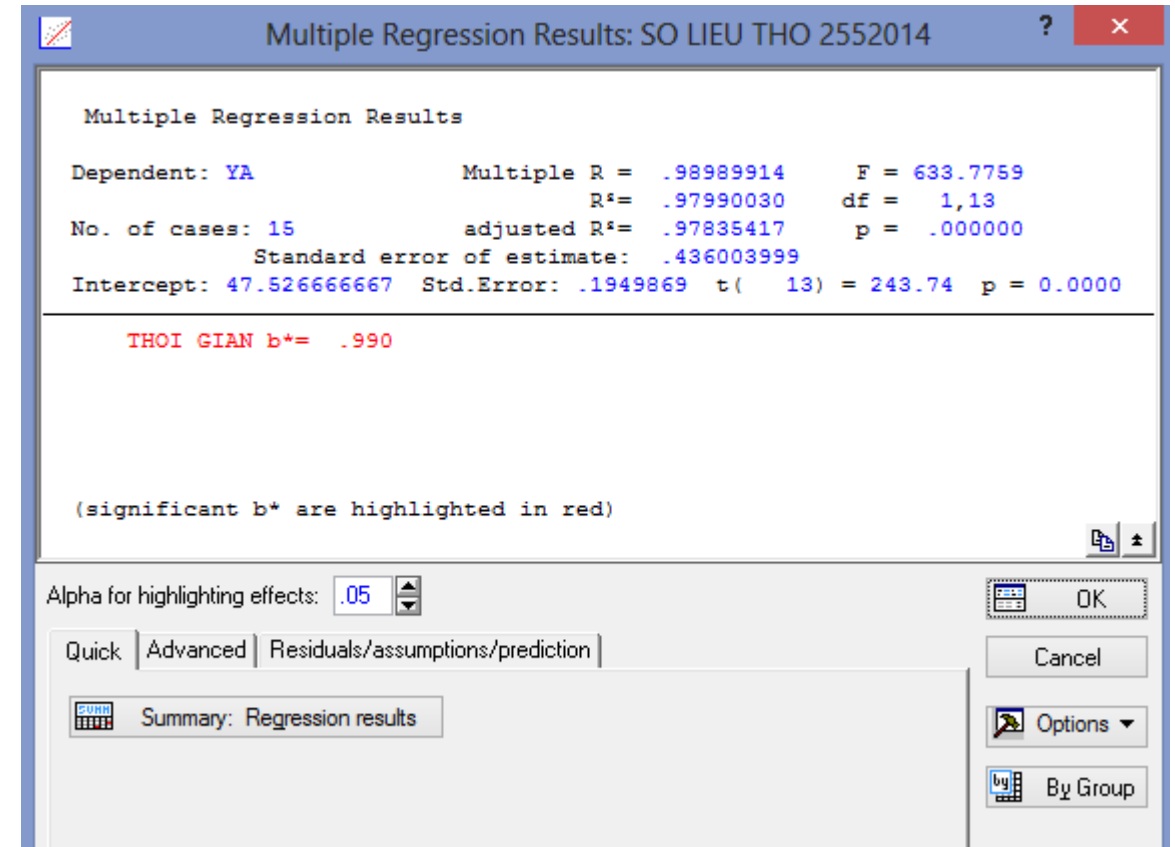
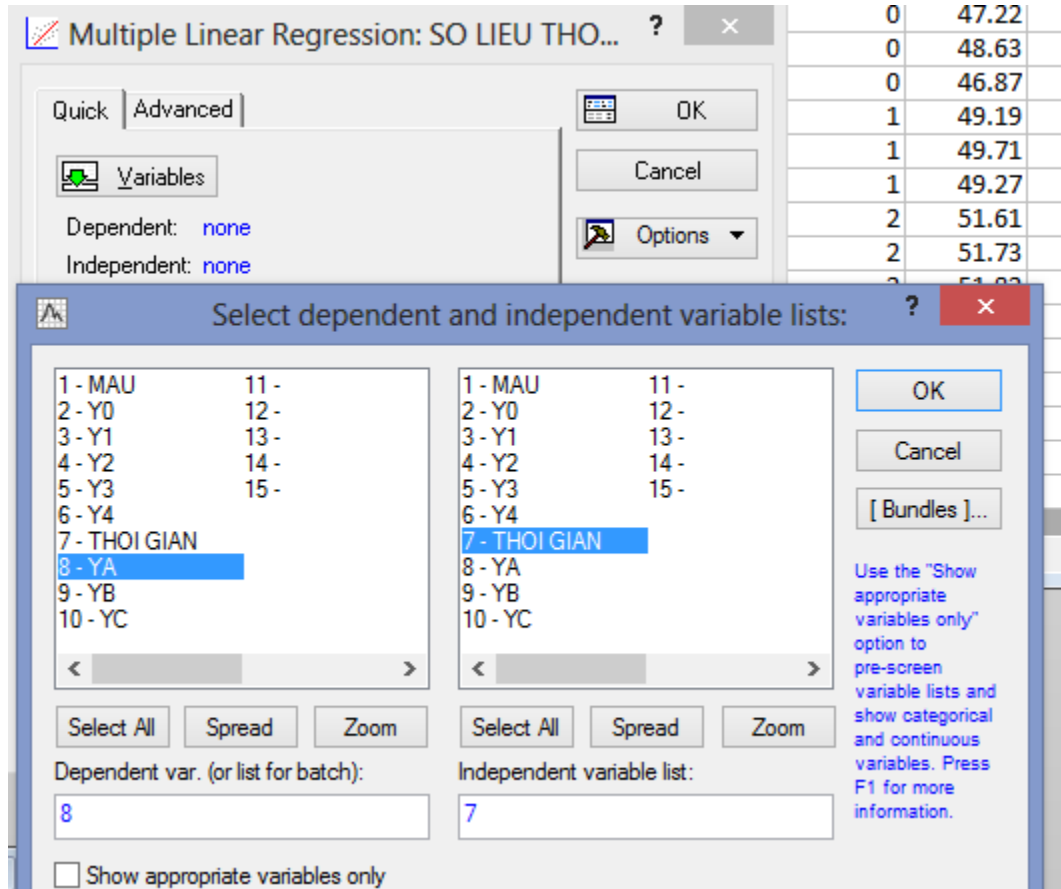
Kết quả thể hiện như bảng bên. Kết quả cho thấy cả 3 mẫu đều có sự khác biệt về chỉ tiêu đánh giá giữa các thời gian bảo quản khác nhau

Tukey HSD test; Variable: YA (Spreadsheet1)					
Marked differences are significant at $p < .05000$					
THOI GIAN	{1}	{2}	{3}	{4}	{5}
	M=47.573	M=49.390	M=51.720	M=53.403	M=55.587
0 {1}		0.005973	0.000177	0.000176	0.000176
1 {2}	0.005973		0.001061	0.000178	0.000176
2 {3}	0.000177	0.001061		0.009869	0.000179
3 {4}	0.000176	0.000178	0.009869		0.001668
4 {5}	0.000176	0.000176	0.000179	0.001668	

Tukey HSD test; Variable: YB (Spreadsheet1)					
Marked differences are significant at $p < .05000$					
THOI GIAN	{1}	{2}	{3}	{4}	{5}
	M=49.140	M=44.527	M=43.393	M=40.447	M=36.507
0 {1}		0.000176	0.000176	0.000176	0.000176
1 {2}	0.000176		0.004777	0.000176	0.000176
2 {3}	0.000176	0.004777		0.000176	0.000176
3 {4}	0.000176	0.000176	0.000176		0.000176
4 {5}	0.000176	0.000176	0.000176	0.000176	

Tukey HSD test; Variable: YC (Spreadsheet1)					
Marked differences are significant at $p < .05000$					
THOI GIAN	{1}	{2}	{3}	{4}	{5}
	M=49.090	M=47.847	M=45.857	M=40.697	M=39.597
0 {1}		0.000683	0.000176	0.000176	0.000176
1 {2}	0.000683		0.000178	0.000176	0.000176
2 {3}	0.000176	0.000178		0.000176	0.000176
3 {4}	0.000176	0.000176	0.000176		0.001607
4 {5}	0.000176	0.000176	0.000176	0.001607	

Để đánh giá sự tương quan, chọn Multiple Linear Regression – Variables – biến phụ thuộc (chọn YA), biến độc lập chọn THOI GIAN- OK-OK



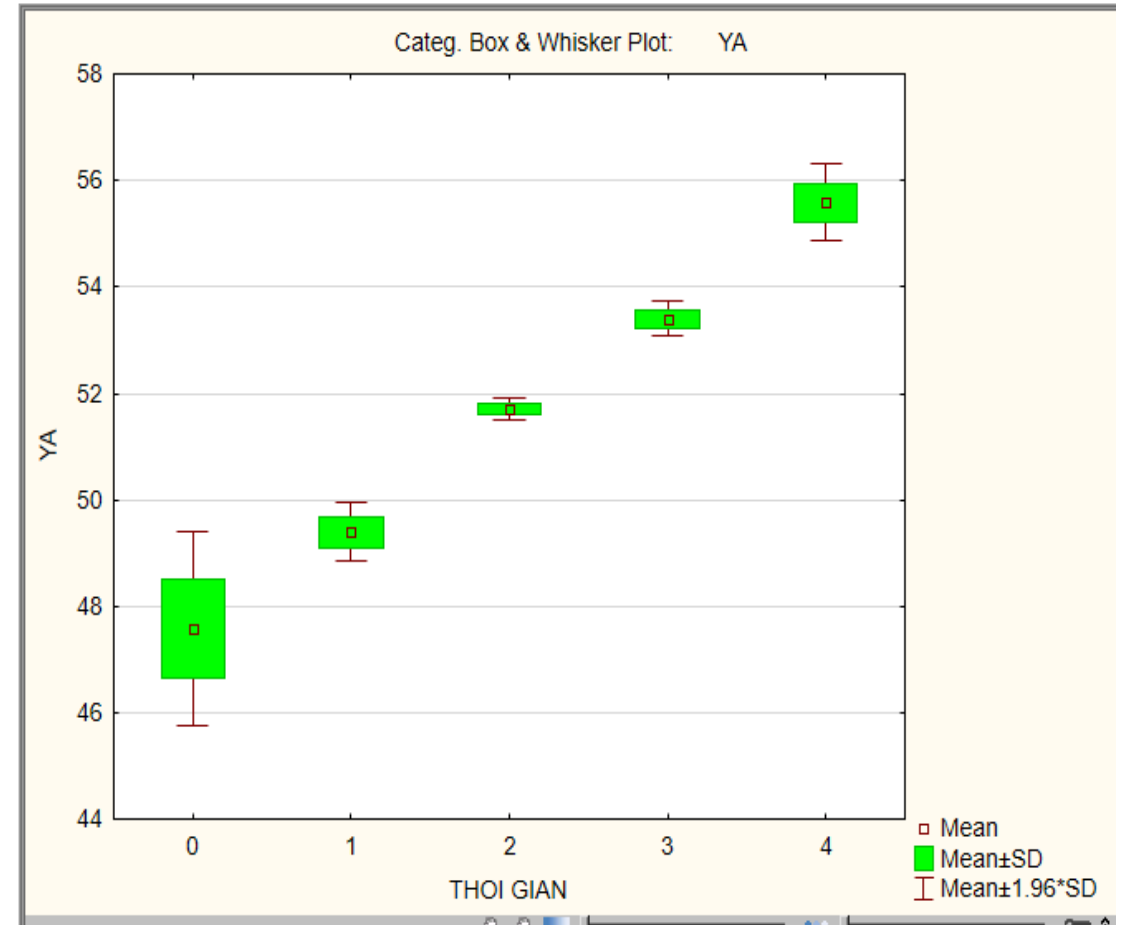
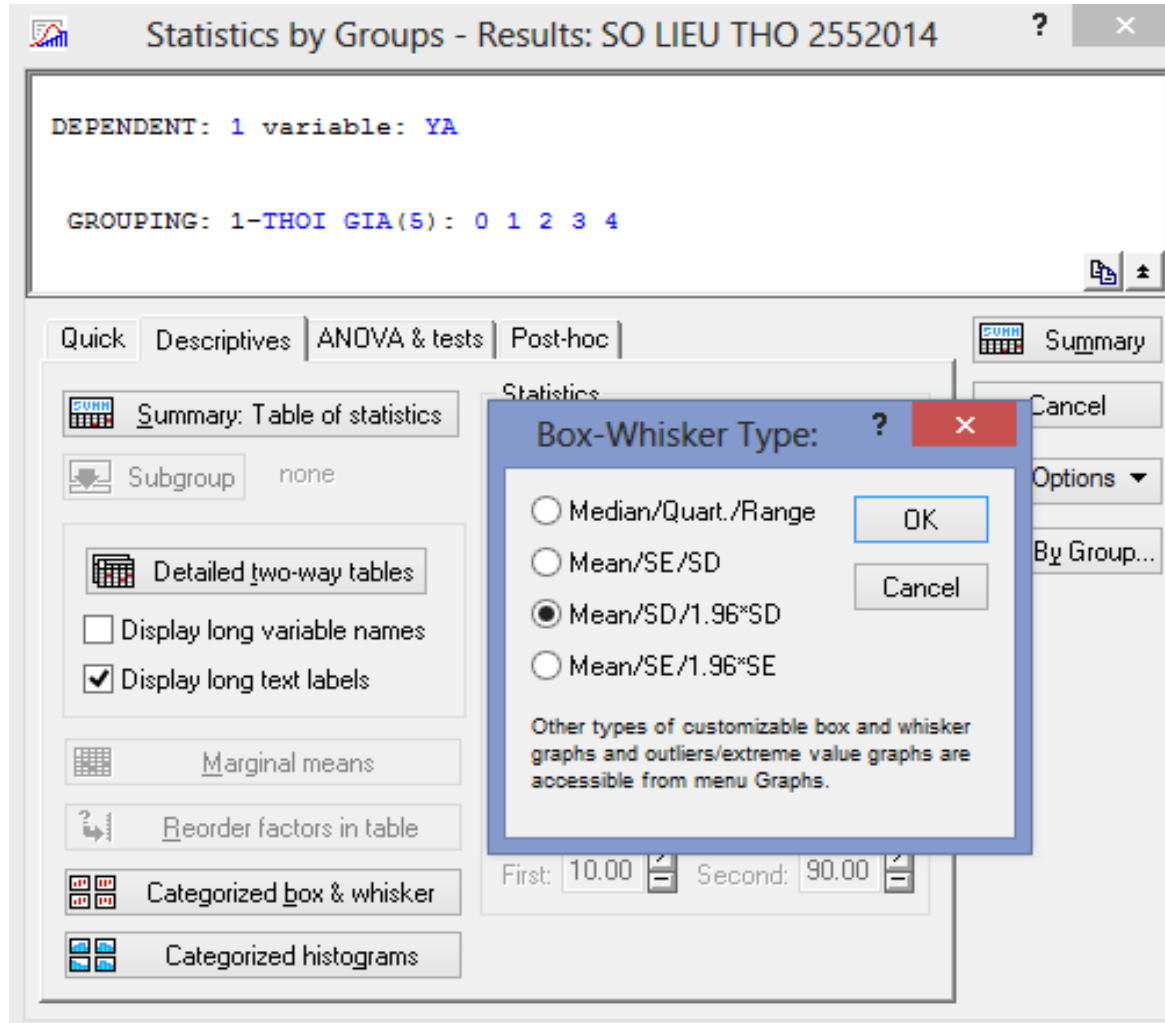
Hoặc chọn Summary: Regression Result

Regression Summary for Dependent Variable: YA (Spreadsheet1)						
R= .98989914 R ² = .97990030 Adjusted R ² = .97835417						
F(1,13)=633.78 p<.00000 Std.Error of estimate: .43600						
N=15	b*	Std.Err. of b*	b	Std.Err. of b	t(13)	p-value
Intercept			47.52667	0.194987	243.7429	0.000000
THOI GIAN	0.989899	0.039321	2.00400	0.079603	25.1749	0.000000

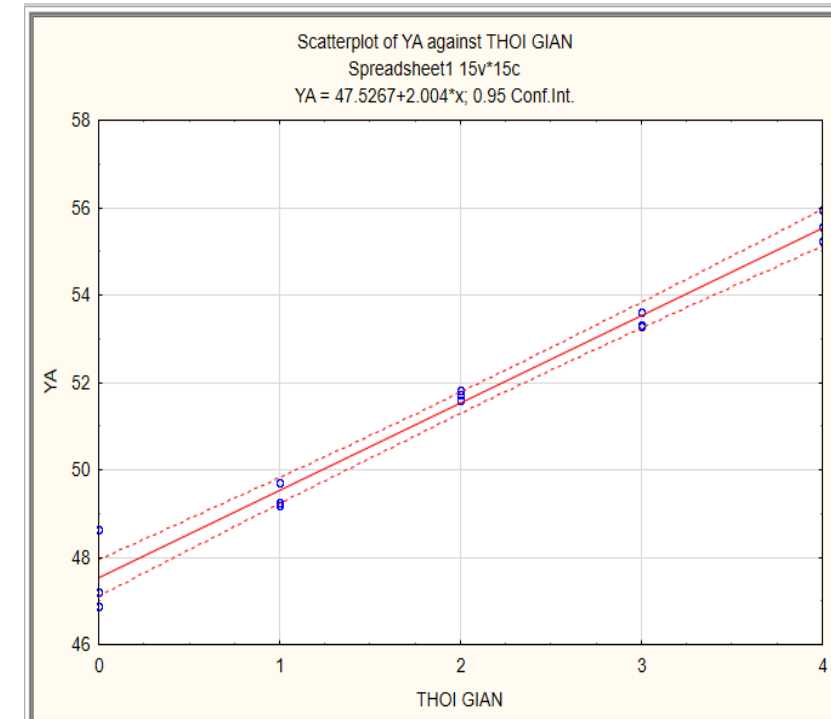
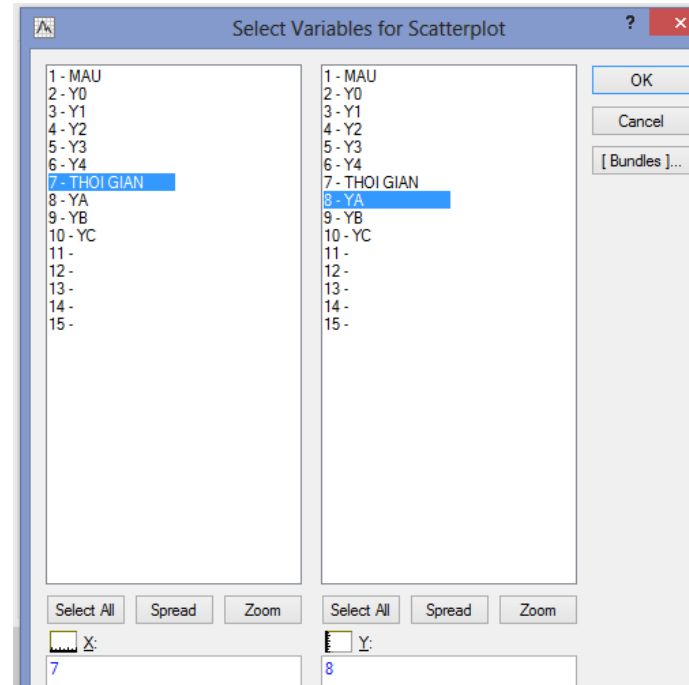
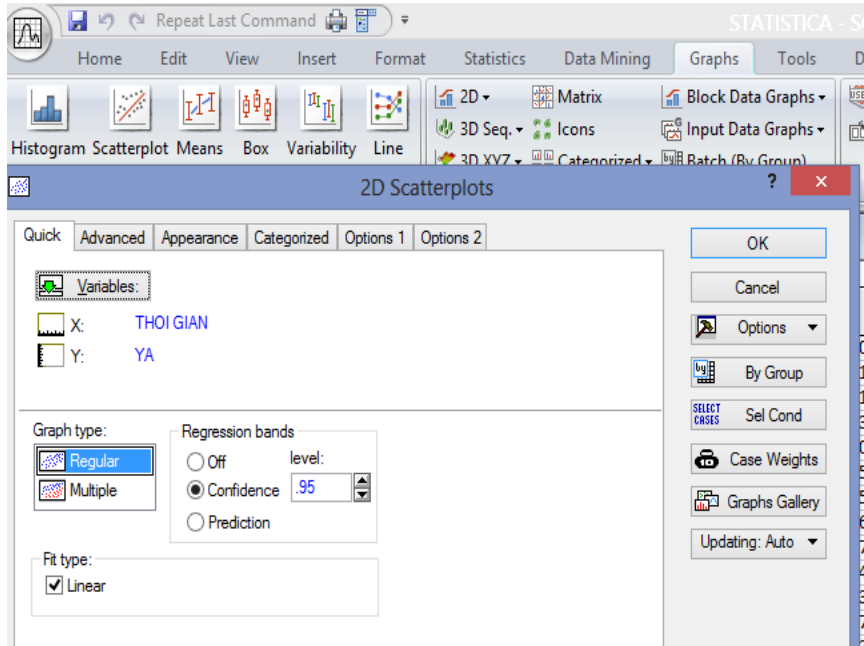
Kết luận: Theo thời gian chỉ tiêu đánh giá của mẫu A có xu hướng tăng (R= 0,99; p<0,0001)

Để vẽ biểu đồ trên STATISTICA

Trong hộp Statistics by Groups chọn Descriptives – Categorized box & whisker – chọn dạng biểu đồ loại gì (trong hộp Box-Whisker)- thường chọn Mean/SD/1.96*SD

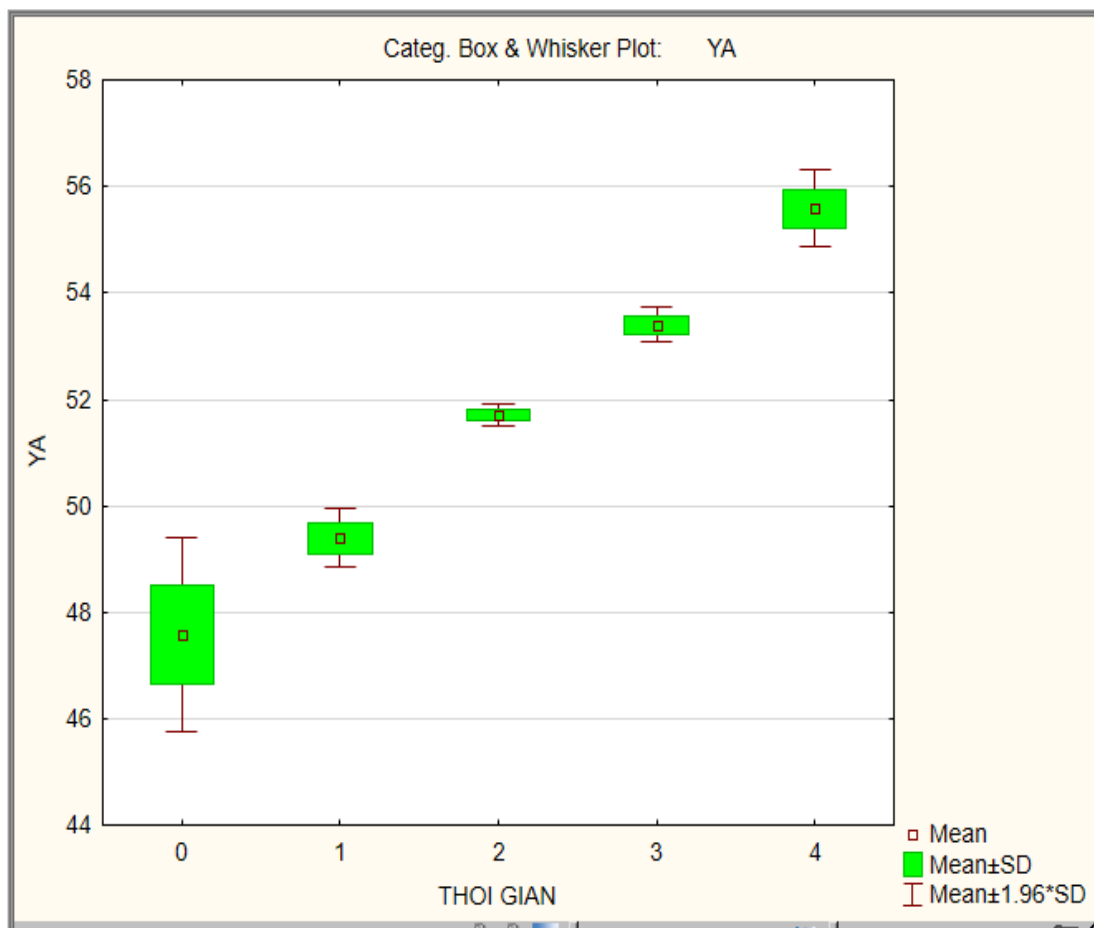


Chọn Graphs – Scatterplot-OK- chọn trục X, Y- OK

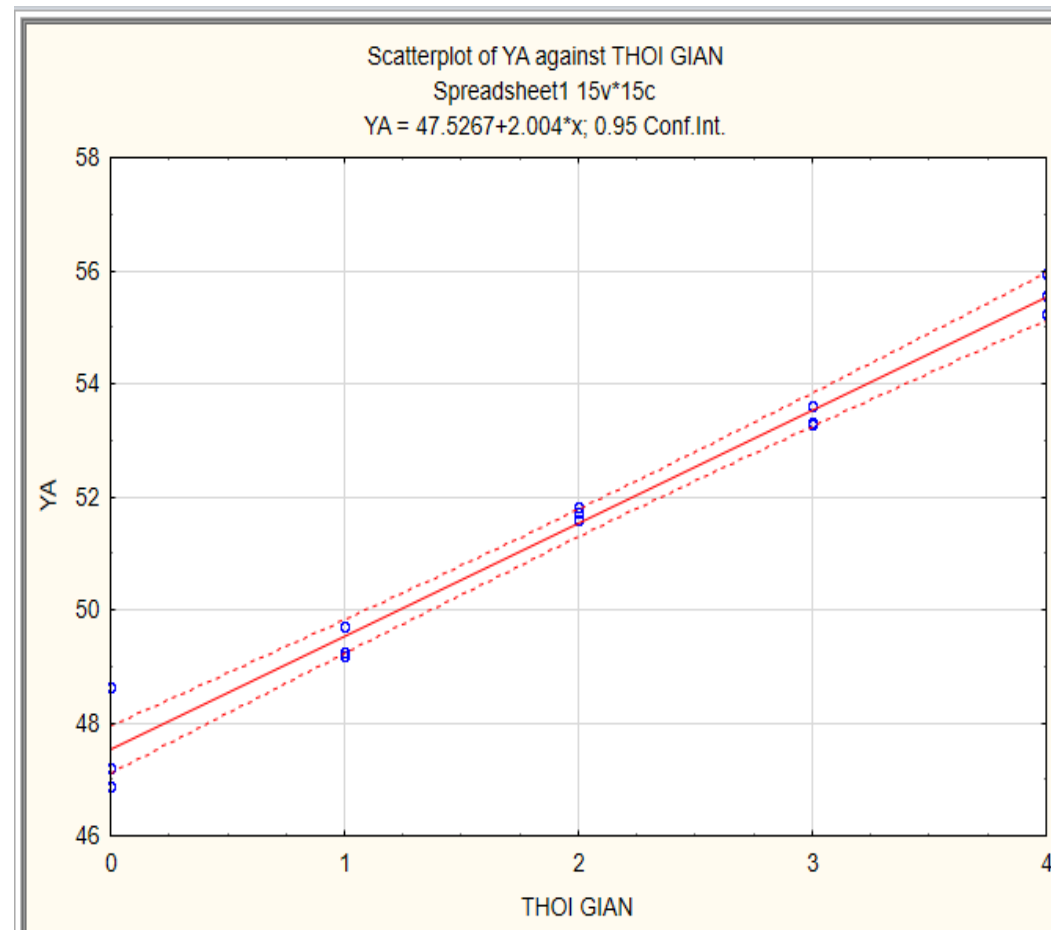


Kết quả về sự khác biệt giữa giá trị trung bình của chỉ tiêu đánh giá thuộc nhóm A theo thời gian bảo quản và mối tương quan giữa chỉ tiêu đánh giá với thời gian bảo quản được thể hiện ở hình a và b (sử dụng STATISTICA)

a



b



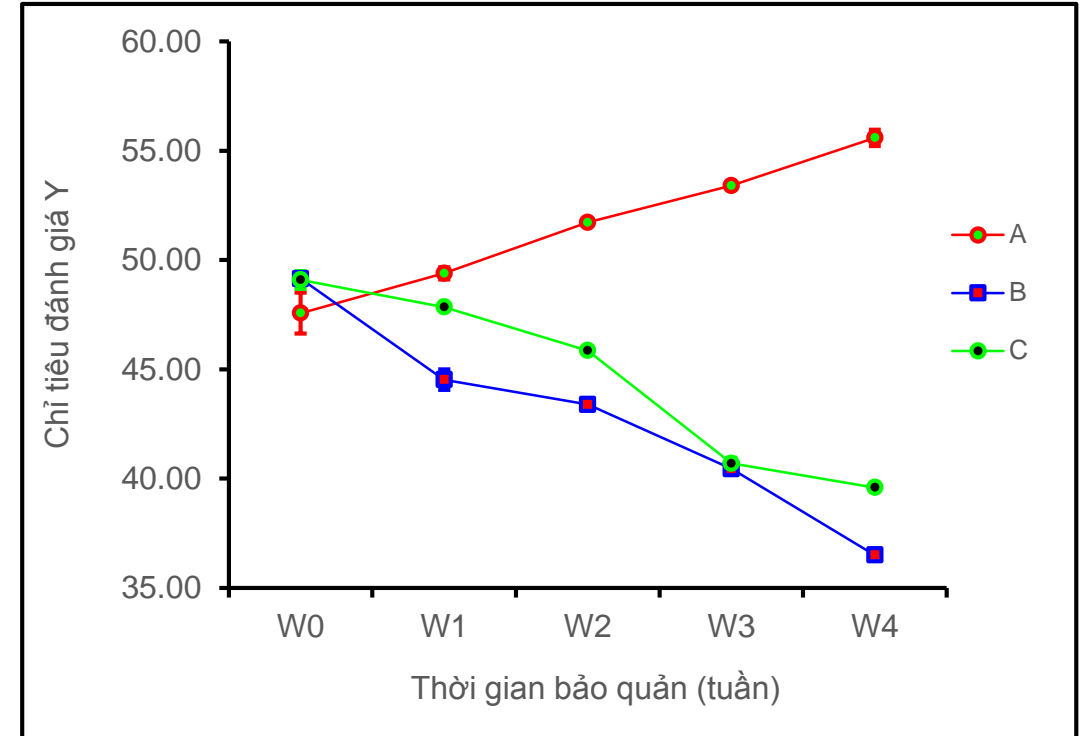
KẾT QUẢ THỂ HIỆN CHO TỪNG NHÓM MẪU

TỔNG HỢP KẾT QUẢ (Sử dụng EXCEL)

TRÌNH BÀY KẾT QUẢ DƯỚI DẠNG BẢNG				
		Mẫu		
		A	B	C
	0	47,57±0,93 ^{xa}	49,14±0,23 ^{xbc}	49,09±0,4 ^{xbc}
Thời gian (tuần)	1	49,39±0,28 ^{ya}	44,53±0,46 ^{yb}	47,85±0,09 ^{yc}
	2	51,72±0,11 ^{za}	43,39±0,16 ^{zb}	45,86±0,11 ^{zc}
	3	53,4±0,17 ^{ha}	40,45±0,25 ^{hbc}	40,7±0,26 ^{hbc}
	4	55,59±0,37 ^{ka}	36,51±0,22 ^{kb}	39,6±0,18 ^{kc}

Các chữ cái x,y,z,h,k trong cùng một cột chỉ sự khác biệt có ý nghĩa thống kê với $p < 0.05$

Các chữ cái a,b,c trong cùng một hàng chỉ sự khác biệt có ý nghĩa thống kê với $p < 0.05$



- Vì độ lệch chuẩn quá bé nên không thể hiện rõ trên biểu đồ
- Với đồ thị đường gấp khúc có thể sử dụng các chữ cái a,b,c để chỉ sự khác biệt hoặc không thể hiện nhưng khi bình luận cần giải thích rõ.