

PHÂN PHỐI F VÀ ỨNG DỤNG TRONG PHÂN TÍCH PHƯƠNG SAI

14.1 Phân phối F:

Nếu gọi U_1 là phân phối khi bình phương với bậc tự do d_1

Và U_2 là phân phối khi bình phương với bậc tự do d_2

Thì phân phối F là tỉ số giữa 2 phân phối khi bình phương U_1 và U_2

$$F \sim \frac{\frac{U_1}{d_1}}{\frac{U_2}{d_2}}$$

Như vậy phân phối F cũng là phân phối khi bình phương có trị trung bình μ và phương sai σ^2

Trị trung bình của phân phối F: (Với $d_2 > 2$)

$$\mu = \frac{d_2}{d_2 - 2}$$

và phương sai là: (với $d_2 > 4$)

$$\sigma^2 = \frac{2d_2^2(d_1 + d_2) - 2}{d_1(d_2 - 2)^2(d_2 - 4)}$$

14.2 Ứng dụng trong phân tích phương sai (ANOVA)

Ví dụ1 : Số lượng tiểu cầu giảm nhiều trong bệnh Sốt xuất huyết (SXH), giảm trung bình trong Sốt dengue (SD) và giảm ít trong Nhiễm siêu vi khác (NSV). Xét nghiệm tiểu cầu 10 bệnh nhân cho mỗi nhóm. Kết quả được ghi nhận trong bảng 14.1. Hỏi có sự khác biệt trị số TC giữa 3 nhóm bệnh?

Bảng 14.1 Trị số tiểu cầu ($\times 1000/\text{mm}^3$) ở bệnh nhân mắc SD, SXH và NSV

SD (n1=10)	SXH (n2=10)	NSV (n3=10)
150	100	140
140	130	200
170	80	210
160	70	230
90	40	100

240	30	120
100	120	220
140	130	160
120	20	150
90	80	170

$\mu_1 = 140$	$\mu_2 = 80$	$\mu_3 = 170$
---------------	--------------	---------------

$s_1^2 = 18400$	$s_2^2 = 14800$	$s_3^2 = 17400$
-----------------	-----------------	-----------------

$10(140-130)^2$	$10(80-130)^2$	$10(170-130)^2$
-----------------	----------------	-----------------

Giả thuyết H_0 : $\mu_1 = \mu_2 = \mu_3$

Giả thuyết H_a : $\mu_1 \neq \mu_2 \neq \mu_3$

Gọi n_1, n_2, n_3 là số ca mỗi nhóm tương ứng

Gọi μ_1, μ_2, μ_3 là trị trung bình của mỗi nhóm tương ứng

Gọi s_1^2, s_2^2, s_3^2 là phương sai của mỗi nhóm tương ứng

Ta có:

$$\mu_1 = \frac{150+140+\dots+90}{10} = 140$$

$$\mu_2 = \frac{100+130+\dots+80}{10} = 80$$

$$\mu_3 = \frac{140+200+\dots+170}{10} = 170$$

Gọi μ là trung bình toàn bộ mẫu:

$$\mu = \frac{140+80+170}{3} = 130$$

Phương sai của nhóm 1 (Sốt dengue)

$$s_1^2 = (150-140)^2 + (140-140)^2 + \dots + (90-140)^2 = 18400$$

Phương sai của nhóm 2 (SXH)

$$s_2^2 = (100-80)^2 + (130-80)^2 + \dots + (80-80)^2 = 14800$$

Phương sai của nhóm 3 (NSV)

$$s_3^2 = (140-170)^2 + (200-170)^2 + \dots + (170-170)^2 = 17400$$

Tổng bình phương phản ánh độ dao động **trong nội bộ** nhóm (**Within-groups** sum of squares):

$$SSW = s_1^2 + s_2^2 + s_3^2 = 50600$$

Tổng bình phương phản ánh độ khác nhau **giữa các nhóm** (**Between-groups** sum of squares):

$$SSB = n_1(\mu_1 - \mu)^2 + n_2(\mu_2 - \mu)^2 + n_3(\mu_3 - \mu)^2$$

$$SSB = 10(140 - 130)^2 + 10(80 - 130)^2 + 10(170 - 130)^2 = 42000$$

Bậc tự do giữa các nhóm = $k - 1 = 2$

Bậc tự do toàn bộ mẫu = $n - k = 27$

Trung bình bình phương giữa các nhóm (between-groups mean squares)

$$MSB = \frac{SSB}{k - 1} = \frac{42000}{(3 - 1)} = 21000$$

Trung bình bình phương trong nội bộ nhóm (within-groups mean squares)

$$MSW = \frac{SSW}{n - k} = \frac{50660}{(30 - 3)} = 1874$$

$$\text{Phép kiểm } F = \frac{MSB}{MSW} = \frac{21000}{1874} = 11.2$$

Đọc bảng phân phối F với bậc tự do của tử số là 2 và mẫu số là 27 ta có

$\alpha_{0.05} = 3.35$. Như vậy $F = 11.2 > 3.45$

Kết luận: bác bỏ $H_0 \rightarrow$ Có sự khác biệt trị tiểu cầu giữa 3 nhóm bệnh (SD, SXH, NSV)

Như vậy ta thấy rằng nếu sự chênh lệch trung bình giữa (**Between**) 3 nhóm càng cao so với chênh lệch trong nội bộ (**Within**) của từng nhóm thì F càng lớn và sự khác biệt càng rõ ràng.

14.2.3 Phân tích phương sai một chiều (one-way analysis of variance) trong SPSS:

Phân tích phương sai 1 chiều khi chỉ có 1 yếu tố (là nhóm bệnh trong ví dụ này)

Nhập số liệu vào SPSS: gồm 2 cột: Nhombenh (1=SD); (2=SXH); (3=NSV) và tieucau $\times 1000/\text{mm}^3$.

	Nhombenh	Tieucau
1	1	150
2	1	140
3	1	170
4	1	160
5	1	90
6	1	240
7	1	100
8	1	140
9	1	120
10	1	90
11	2	100
12	2	130
13	2	80
14	2	70
15	2	40
16	2	30
17	2	120
18	2	130
19	2	20
20	2	80
21	3	140
22	3	200
23	3	210
24	3	230
25	3	100
26	3	120
27	3	220
28	3	160
29	3	150
30	3	170

Vào menu: **Analyze > Compare means > One-way ANOVA**

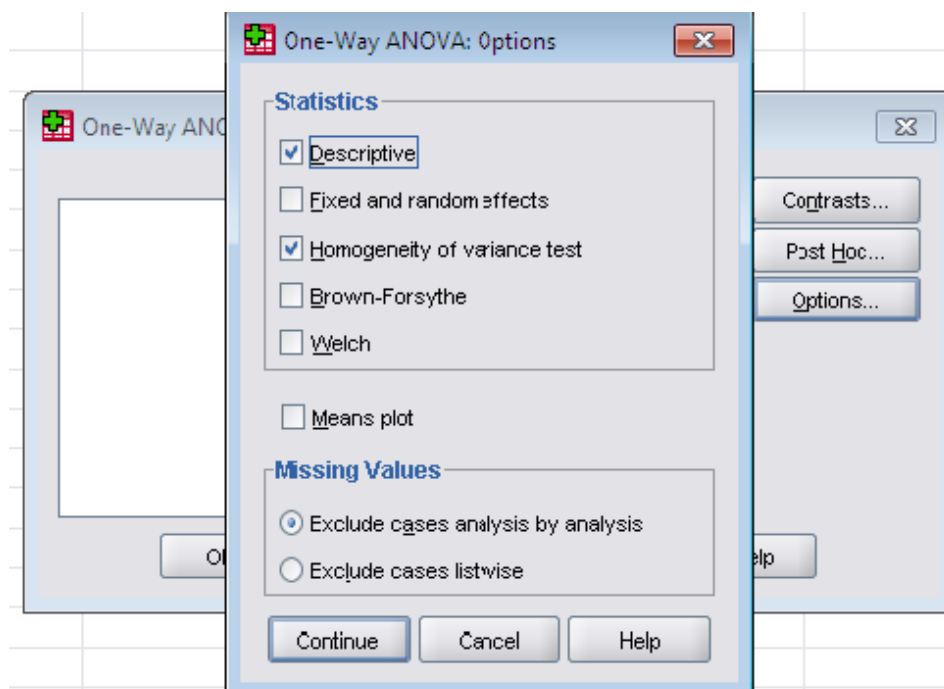
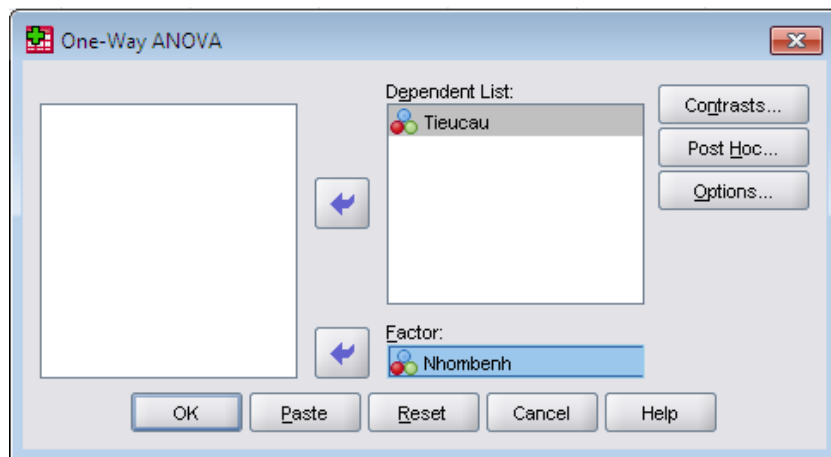
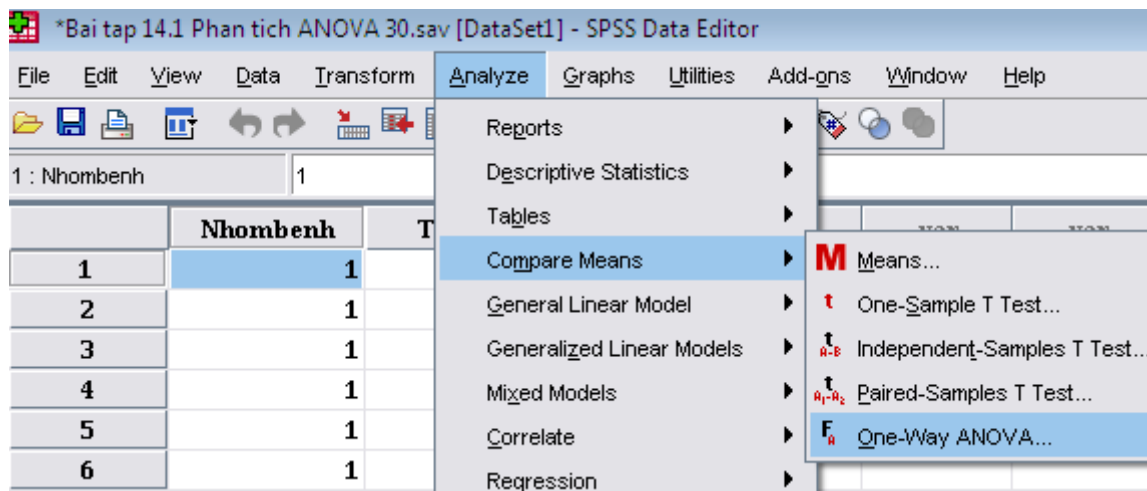
Chọn tieucau trong ô bên trái, nhấp mũi tên, chuyển vào ô Dependent List

Chọn Nhombenh, chuyển vào ô Factor

Nhấn Options: Nhấp dấu nháy ✓ vào ô Descriptive và ô Homogeneity of variance test

Nhấp continue

Nhấn OK (như hình sau)



Kết quả:

	N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean	
					Lower Bound	Upper Bound
1	10	140.0000	45.21553	14.29841	107.6548	172.3452
2	10	80.0000	40.55175	12.82359	50.9910	109.0090
3	10	170.0000	43.96969	13.90444	138.5460	201.4540
Total	30	130.0000	56.50755	10.31682	108.8997	151.1003

Test of Homogeneity of Variances

Tieu cau

Levene Statistic	df1	df2	Sig.
.084	2	27	.920

ANOVA

Tieu cau

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	42000.000	2	21000.000	11.206	.000
Within Groups	50600.000	27	1874.074		
Total	92600.000	29			

Trị số trung bình tiểu cầu (mean) và độ lệch chuẩn (Std. Deviation) của từng nhóm:
SD (1): N=10, trung bình= 140 (ĐLC 45); SXH (2): N=10, trung bình= 80 (ĐLC 40) và

NSV(3): N=10, trung bình= 170 (ĐLC 56)

Levene Statistic=0,084, và Sig.= 0,920 cho biết không có sự sai biệt nhiều về phương sai giữa 3 nhóm hoặc có tính đồng nhất (homogeneity) nên có thể dùng kiểm định ANOVA.

Tổng bình phương **giữa** các nhóm (Sum of squares- Between groups)=42000

Bậc tự do (df)=2 (3 nhóm -1)

Bình phương trung bình (mean Square)=SSB/2= 21000

Tổng bình phương toàn bộ mẫu (Sum of squares- Within groups)=506000

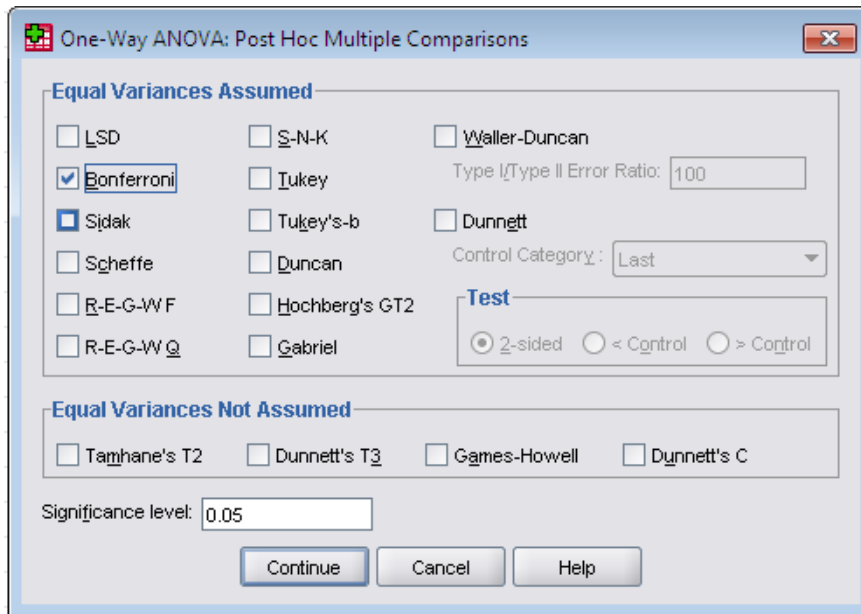
Bậc tự do (df)=30-3=27

Bình phương trung bình (mean Square)=50600/27=1874

F= 21000/1874=11,206, Ý nghĩa TK (Sig.) với p=0,000

Làm tiếp thủ tục Post Hoc để xác định sự khác nhau về trung bình của từng cặp nhóm, thông thường dùng kiểm định Bonferroni trong Post Hoc

Nhấp Post Hoc → Nhấp dấu nháy ✓ vào ô Bonfferoni → Nhấp Continue và nhấn OK.



Post Hoc Tests

Multiple Comparisons

Dependent Variable: TC

Bonferroni

(I) nhóm	(J) nhóm	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
1	2	60.00000*	19.36013	.013	10.5841	109.4159
	3	-30.00000	19.36013	.399	-79.4159	19.4159
2	1	-60.00000*	19.36013	.013	-109.4159	-10.5841
	3	-90.00000*	19.36013	.000	-139.4159	-40.5841
3	1	30.00000	19.36013	.399	-19.4159	79.4159
	2	90.00000*	19.36013	.000	40.5841	139.4159

*. The mean difference is significant at the .05 level.

Sai biệt trung bình giữa nhóm 1 (SD) và nhóm 2 (SXH); 60.0, có ý nghĩa thống kê với $p=0.013$

Sai biệt trung bình giữa nhóm 1 (SD) và nhóm 3 (NSV); -30.0, không có ý nghĩa thống kê với $p=0.399$

Sai biệt trung bình giữa nhóm 2 (SXH) và nhóm 3 (NSV); -90.0, có ý nghĩa thống kê với $p=0.000$

Tóm lại: trị tiêu cầu của SXH thấp hơn SD và NSV có ý nghĩa thống kê, nhưng không có sự khác biệt trị tiêu cầu giữa 2 nhóm SD và NSV.

14.2.4 Phân tích phương sai hai chiều (two-way analysis of variance) trong SPSS:

Trong ví dụ trên, giả sử rằng làm xét nghiệm tiêu cầu sớm (≤ 2 ngày của bệnh) thì mức độ giảm tiêu cầu của nhóm SXH không khác biệt so với nhóm SD hoặc NSV, như vậy ngoài yếu tố nhóm bệnh, ta có thể 1 yếu tố (điều kiện) thứ hai là xét nghiệm máu sớm “1” hoặc trễ “0”. Như vậy ta có 3 nhóm bệnh (SD, SXH, NSV) và 2 điều kiện (xét nghiệm sớm và trễ)

	Nhóm bệnh		
Điều kiện	SD=1	SXH=2	NSV=3
Sớm=1	100, 120, 140, 140, 240	80, 100, 120 , 130, 130	100, 160, 200, 220, 230
Trễ=0	90, 90, 150, 160, 170	20, 30, 40, 70, 80	120, 140, 150, 170, 210

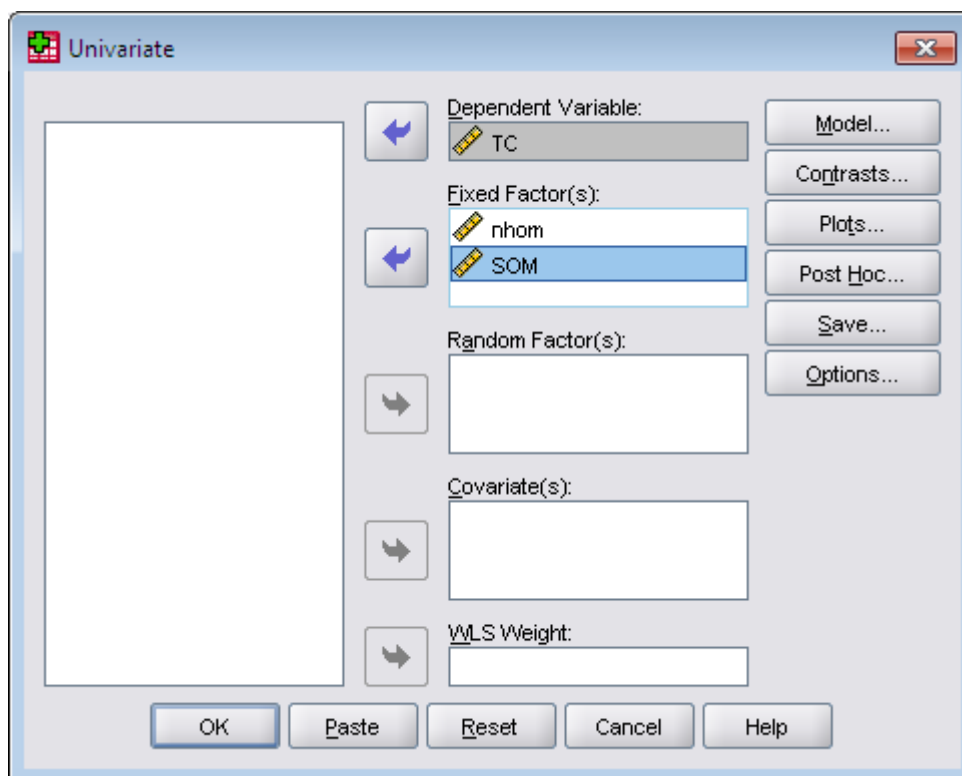
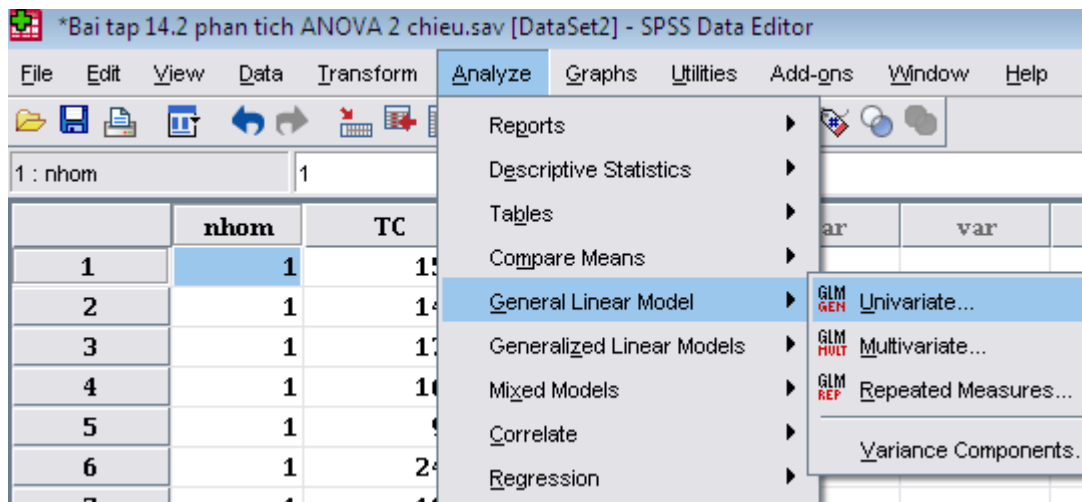
Nhập số liệu vào SPSS, gồm 4 cột:
 Nhóm “1” (sốt dengue), “2” sốt xuất huyết, “3” Nhiễm siêu vi
 TC (số lượng tiêu cầu $1000/\text{mm}^3$)

Som; “1” xét nghiệm sớm

“0” xét nghiệm trễ

	nhom	TC	SOM
1	1	150	0
2	1	140	1
3	1	170	0
4	1	160	0
5	1	90	0
6	1	240	1
7	1	100	1
8	1	140	1
9	1	120	1
10	1	90	0
11	2	100	1
12	2	130	1
13	2	80	0
14	2	70	0
15	2	40	0
16	2	30	0
17	2	120	1
18	2	130	1
19	2	20	0
20	2	80	1
21	3	140	0
22	3	200	1
23	3	210	0
24	3	230	1
25	3	100	1
26	3	120	0
27	3	220	1
28	3	160	1
29	3	150	0
30	3	170	0

Vào menu: **Analyze**> **General linear model**> **Univariate** như hình sau:



Nhập TC vào ô Dependent Variable, nhập nhom và som vào ô Fixed Factor(s)
 Nhấp OK. Kết quả phân tích phương sai hai yếu tố trong SPSS như sau:

→ Univariate Analysis of Variance

Tests of Between-Subjects Effects

Dependent Variable: TC

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	54320.000 ^a	5	10864.000	6.811	.000
Intercept	507000.000	1	507000.000	317.868	.000
nhom	42000.000	2	21000.000	13.166	.000
SOM	9013.333	1	9013.333	5.651	.026
nhom * SOM	3306.667	2	1653.333	1.037	.370
Error	38280.000	24	1595.000		
Total	599600.000	30			
Corrected Total	92600.000	29			

a. R Squared = .587 (Adjusted R Squared = .500)

Kết quả phân tích trên có thể tóm tắt trong bảng sau:

Nguồn biến thiên (Source)	Tổng bình phương (type III)	Bậc tự do (df)	Trung bình bình phương	Kiểm định F	Ý nghĩa TK
Khác biệt giữa 3 nhóm	42000	2 (3 nhóm -1)	42000/2=21000	21000/1595=13.16	0.000
Sớm	9013.33	1 (2 điều kiện-1)	9013/1=9013	9013/1595=5.65	0.026
Tương tác Nhóm* sớm	3306.66	2	3306.6/2=1653	1653/1595=1.03	0.370
Phần dư (Error)	38280,000	24 (N- điều kiện (2)x nhóm (3))	38280/24=1595		
Tổng cộng	599600,000	30			

Kết luận: Có sự khác biệt tiểu cầu giữa 3 nhóm (SD, SXH, NSV) với $p=0.000$

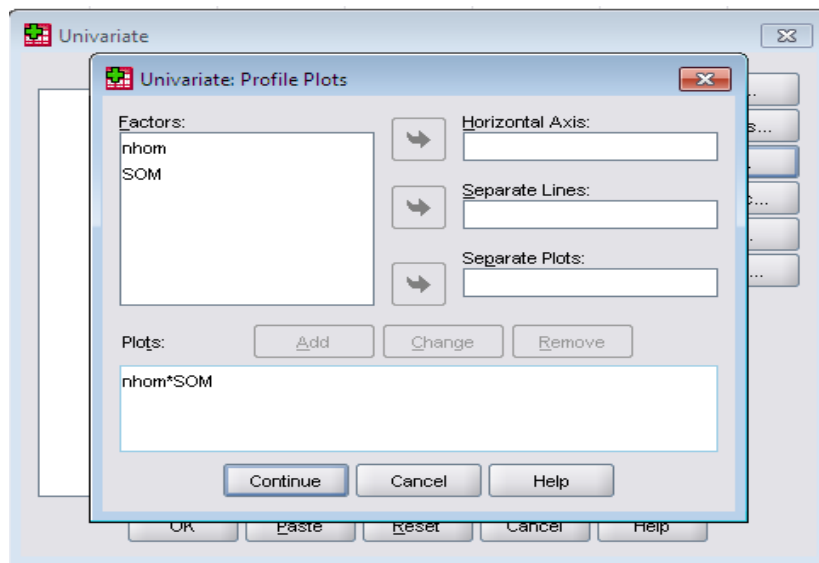
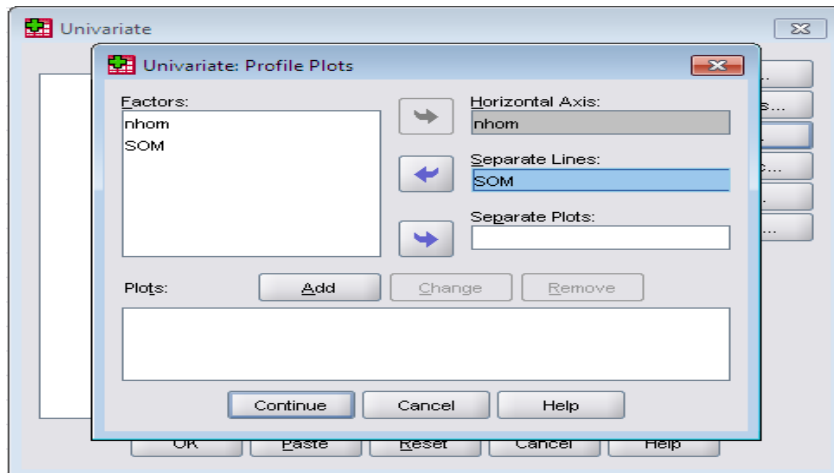
Có sự khác biệt tiểu cầu giữa 2 điều kiện thử sớm hoặc trễ ($p=0.026$)

Không có sự tương tác giữa nhóm bệnh và điều kiện thử sớm hoặc trễ ($p=0.370$).

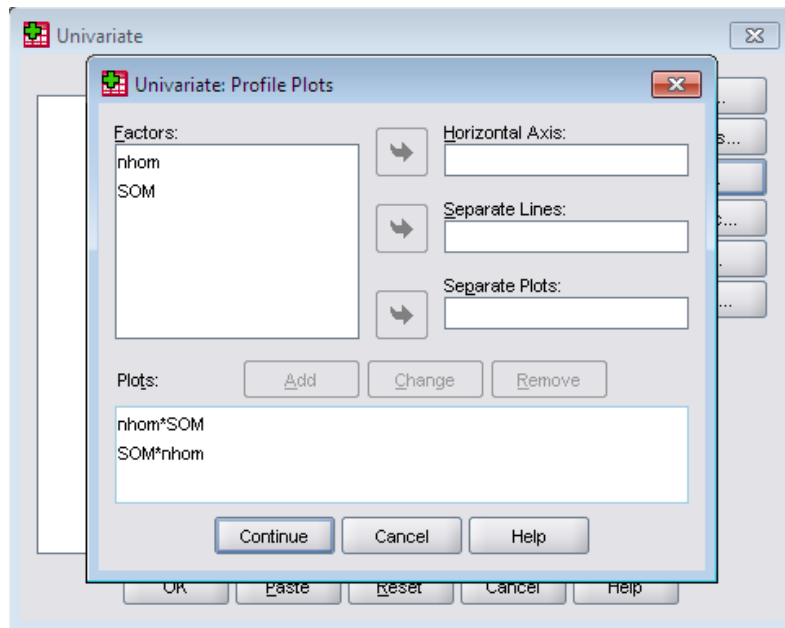
Để xem biểu đồ tương tác, nhấp Plots sẽ xuất hiện màn hình sau.

Nhấp nhóm vào ô Horizontal Axis, sớm vào ô Separate Lines, nhấp Add sẽ thấy xuất

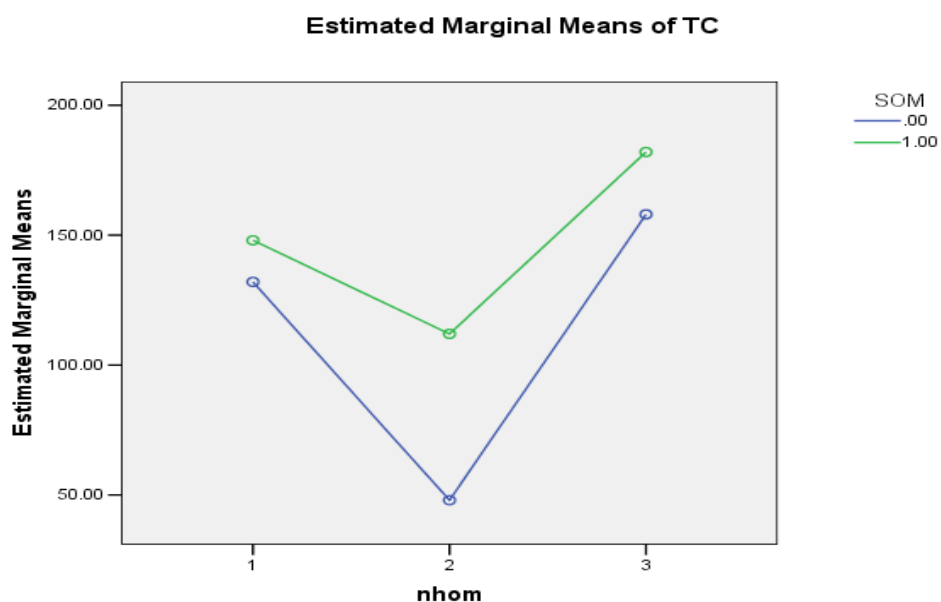
hiện nhóm*SOM trong ô bên dưới



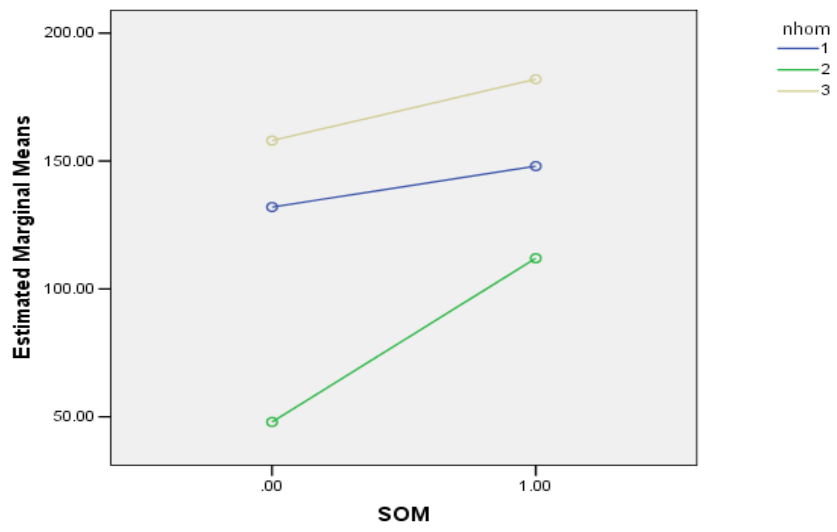
Và cũng tương tự nhấp SOM vào ô Horizontal Axis và nhóm vào ô separate Lines, sau đó nhấn nút Add, sẽ thấy xuất hiện SOM*nhom trong ô bên dưới



Kết quả cho thấy 2 đường gần song song, trị tiêu cầu trung bình thấp nhất trong nhóm SXH và trị tiêu cầu có khuynh hướng giảm đều cả 3 nhóm nếu được làm xét nghiệm trẻ, chứng tỏ **không có sự tương tác** giữa nhóm bệnh và điều kiện thử sớm hoặc trẻ. Nếu có tương tác, thường các đường không song song mà bắt chéo nhau (sẽ trình bày trong bài hiệu ứng tương tác)



Estimated Marginal Means of TC



Phụ lục. Bảng phân phối F (df1: bậc tự do của tử số và df2: bậc tự do của mẫu số)

F - Distribution ($\alpha = 0.05$ in the Right Tail)										
df ₂	df ₁	Numerator Degrees of Freedom								
		1	2	3	4	5	6	7	8	9
Denominator Degrees of Freedom	1	161.45	199.50	215.71	224.58	230.16	233.99	236.77	238.88	240.54
	2	18.513	19.000	19.164	19.247	19.296	19.330	19.353	19.371	19.385
	3	10.128	9.5521	9.2766	9.1172	9.0135	8.9406	8.8867	8.8452	8.8123
	4	7.7086	9.9443	6.5914	6.3882	6.2561	6.1631	6.0942	6.0410	6.9988
	5	6.6079	5.7861	5.4095	5.1922	5.0503	4.9503	4.8759	4.8183	4.7725
	6	5.9874	5.1433	4.7571	4.5337	4.3874	4.2839	4.2067	4.1468	4.0990
	7	5.5914	4.7374	4.3468	4.1203	3.9715	3.8660	3.7870	3.7257	3.6767
	8	5.3177	4.4590	4.0662	3.8379	3.6875	3.5806	3.5005	3.4381	3.3881
	9	5.1174	4.2565	3.8625	3.6331	3.4817	3.3738	3.2927	3.2296	3.1789
	10	4.9646	4.1028	3.7083	3.4780	3.3258	3.2172	3.1355	3.0717	3.0204
	11	4.8443	3.9823	3.5874	3.3567	3.2039	3.0946	3.0123	2.9480	2.8962
	12	4.7472	3.8853	3.4903	3.2592	3.1059	2.9961	2.9134	2.8486	2.7964
	13	4.6672	3.8056	3.4105	3.1791	3.0254	2.9153	2.8321	2.7669	2.7144
	14	4.6001	3.7389	3.3439	3.1122	2.9582	2.8477	2.7642	2.6987	2.6458
	15	4.5431	3.6823	3.2874	3.0556	2.9013	2.7905	2.7066	2.6408	2.5876
	16	4.4940	3.6337	3.2389	3.0069	2.8524	2.7413	2.6572	2.5911	2.5377
	17	4.4513	3.5915	3.1968	2.9647	2.8100	2.6987	2.6143	2.5480	2.4943
	18	4.4139	3.5546	3.1599	2.9277	2.7729	2.6613	2.5767	2.5102	2.4563
	19	4.3807	3.5219	3.1274	2.8951	2.7401	2.6283	2.5435	2.4768	2.4227
	20	4.3512	3.4928	3.0984	2.8661	2.7109	2.5990	2.5140	2.4471	2.3928
	21	4.3248	3.4668	3.0725	2.8401	2.6848	2.5727	2.4876	2.4205	2.3660
	22	4.3009	3.4434	3.0491	2.8167	2.6613	2.5491	2.4638	2.3965	2.3419
	23	4.2793	3.4221	3.0280	2.7955	2.6400	2.5277	2.4422	2.3748	2.3201
	24	4.2597	3.4028	3.0088	2.7763	2.6207	2.5082	2.4226	2.3551	2.3002
	25	4.2417	3.3852	2.9912	2.7587	2.6030	2.4904	2.4047	2.3371	2.2821
	26	4.2252	3.3690	2.9752	2.7426	2.5868	2.4741	2.3883	2.3205	2.2655
	27	4.2100	3.3541	2.9604	2.7278	2.5719	2.4591	2.3732	2.3053	2.2501
	28	4.1960	3.3404	2.9467	2.7141	2.5581	2.4453	2.3593	2.2913	2.2360
	29	4.1830	3.3277	2.9340	2.7014	2.5454	2.4324	2.3463	2.2783	2.2229
	30	4.1709	3.3158	2.9223	2.6896	2.5336	2.4205	2.3343	2.2662	2.2107
40	4.0847	3.2317	2.8387	2.6060	2.4495	2.3359	2.2490	2.1802	2.1240	
60	4.0012	3.1504	2.7581	2.5252	2.3683	2.2541	2.1665	2.0970	2.0401	
120	3.9201	3.0718	2.6802	2.4472	2.2899	2.1750	2.0868	2.0164	1.9588	
∞	3.8415	2.9957	2.6049	2.3719	2.2141	2.0986	2.0096	1.9384	1.8799	