

Phân tích biến lượng

Chương 3

- Mục tiêu của ANOVA
- ANOVA một chiều
- ANOVA hai chiều
- Qui hoạch hình vuông La tin
- Qui hoạch hình vuông La tin- Hy lạp
- Qui hoạch khối La Tin

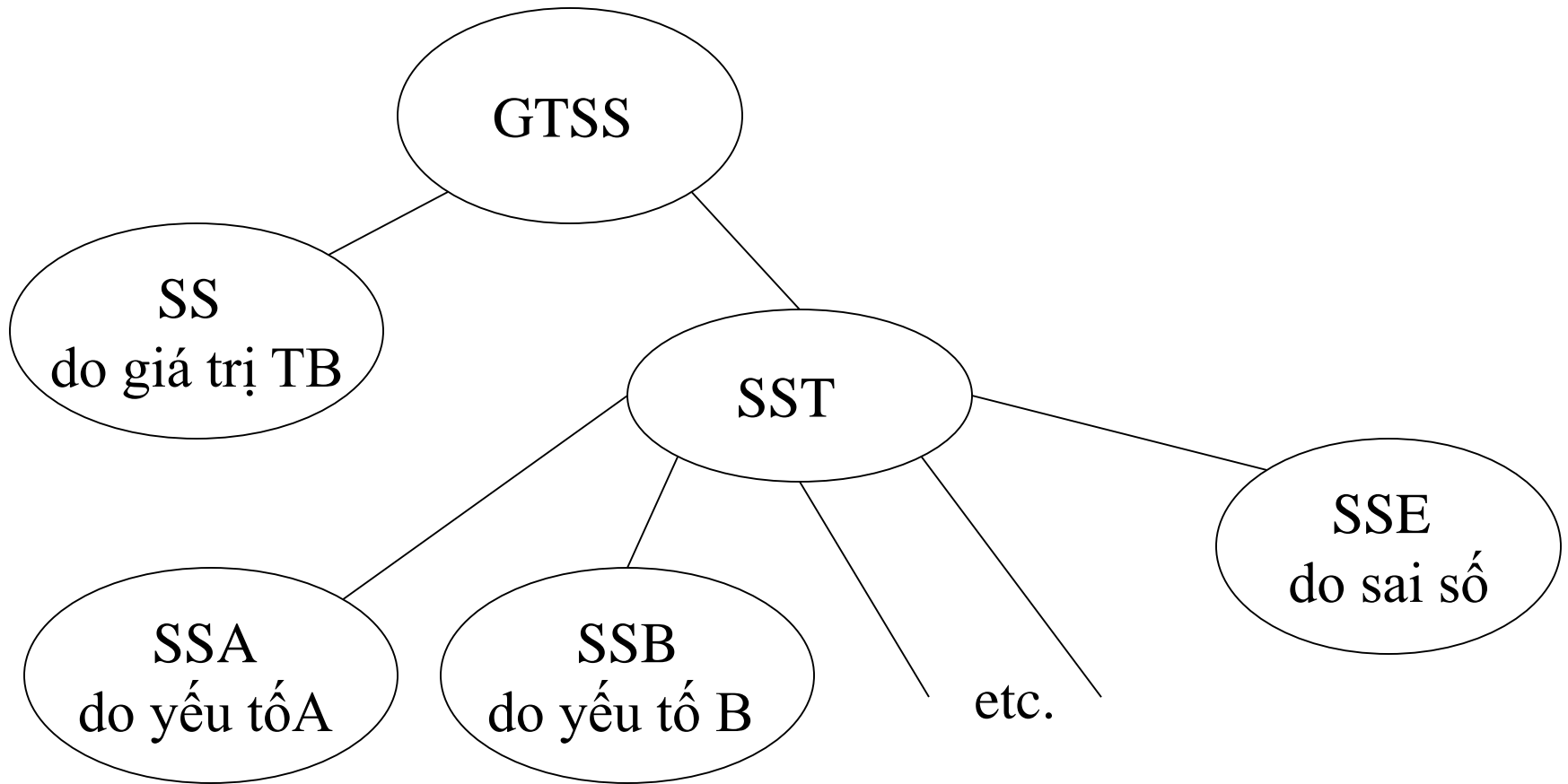
3.1. Mục tiêu của ANOVA

- ANOVA nghiên cứu ảnh hưởng của các yếu tố đến sự thay đổi giá trị của đáp ứng qua việc đánh giá sự thay đổi của giá trị trung bình của chúng
- ANOVA sử dụng tính cộng của biến lượng của các biến ngẫu nhiên
- ANOVA là một công cụ rất mạnh khi khảo sát nhiều yếu tố đồng thời (phù hợp với qui hoạch thực nghiệm)
- Cơ sở của ANOVA là tách biến lượng tổng thành các biến lượng thành phần, mỗi thành phần này tương ứng với một nguồn thay đổi

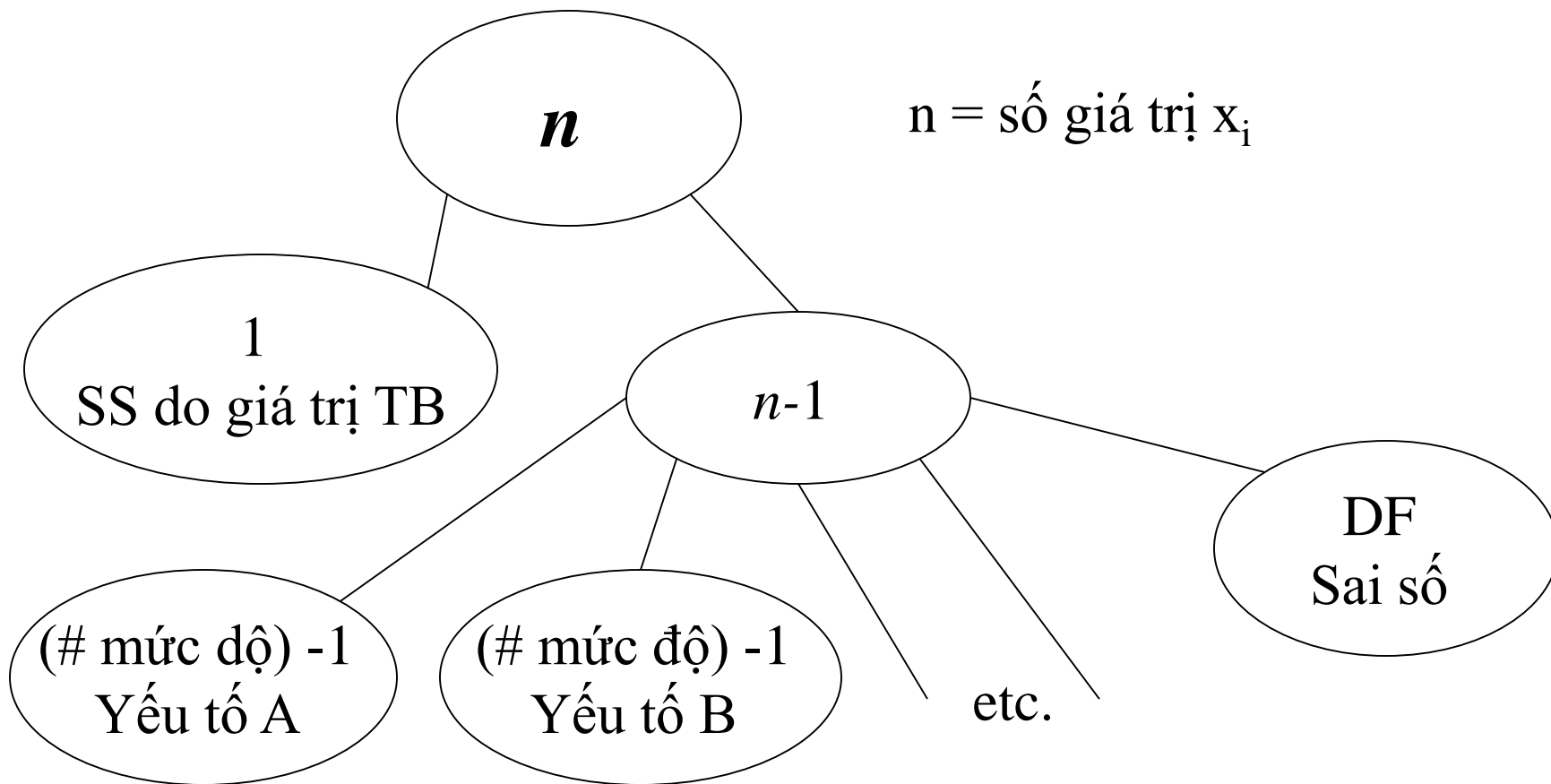
- Biến lượng của mẫu tương ứng sẽ được so sánh với biến lượng do sai số ngẫu nhiên
- Kiểm nghiệm được sử dụng là kiểm nghiệm F
- Tính toán dựa trên các giả thiết
 - Sai số quan sát ngẫu nhiên được phân bố theo hàm phân bố bình thường (hàm phân bố Gauss).
 - Các yếu tố chỉ ảnh hưởng đến sự thay đổi giá trị trung bình. Biến lượng quan sát vẫn không thay đổi.
 - Các thực nghiệm có độ chính xác như nhau.

- Trong ANOVA, biến lượng được tính qua bình phương trung bình (MSS). Bình phương trung bình là tỉ số của tổng bình phương (SS) và độ tự do (DF)
- Có 3 loại tổng bình phương
 - Tổng bình phương chung: SST
 - Tổng bình phương yếu tố: SSA
 - Tổng bình phương sai số : SSE

Các thành phần SS



Các thành phần độ tự do (DF)



Cách tính tổng bình phương

- Tổng bình phương toàn phần $GTSS = \sum_{i=1}^n x_i^2$
- Tổng bình phương do trung bình $SSM = n\mu^2$
- Tổng bình phương chung $SST = \sum_{i=1}^n (x_i - \mu)^2$

- Tổng bình phương do yếu tố

$$SSA = replication \# \left[(m_{A1} - \mu)^2 + (m_{A2} - \mu)^2 + (m_{A3} - \mu)^2 \right]$$

- Tổng bình phương do sai số SSE
 - Bằng 0 nếu không có thí nghiệm lập
 - Ước tính bằng phương pháp gộp (pooling). Gộp các yếu tố có đóng góp thấp nhất vào TSS

F-statistic

- Biến lượng sai số = $\frac{\text{SS của sai số}}{\text{độ tự do của sai số}}$
- $F = \frac{\text{Bình phương trung bình yếu tố}}{\text{Biến lượng sai số}}$
- Bình phương trung bình của yếu tố = $\frac{\text{SS của yếu tố}}{\text{DF của yếu tố}}$
 - $F=1$ ảnh hưởng của yếu tố ngang với sai số
 - $F=2$ ảnh hưởng của yếu tố sát biên
 - $F>4$ ảnh hưởng của yếu tố đáng kể

3.2. ANOVA một chiều

- ANOVA một chiều dùng để kiểm nghiệm sự đồng nhất của hai hay nhiều giá trị trung bình của mẫu thống kê
 - ANOVA một chiều sử dụng kiểm nghiệm F nên thường gọi là ANOVA F
- Đây là sự mở rộng của kiểm nghiệm t đối với 2 mẫu độc lập
- Trường hợp chỉ có 2 nhóm thì kiểm nghiệm t và ANOVA một chiều giống nhau và luôn luôn cho cùng giá trị p
- ANOVA giúp nhà phân tích tránh rủi ro sai số loại I quá lớn khi khảo sát nhiều giá trị trung bình

- Khi so sánh nhiều giá trị trung bình sử dụng kiểm nghiệm t thì phải tiến hành một loạt kiểm nghiệm t (vì kiểm nghiệm t một lần chỉ kiểm nghiệm chỉ 2 giá trị trung bình)
- Mặc dù mỗi kiểm nghiệm chỉ thực hiện với một mức ý nghĩa α , nhưng mức ý nghĩa sẽ tích lũy theo loạt kiểm nghiệm do đó ở kiểm nghiệm cuối cùng sẽ có mức ý nghĩa rất lớn

- ANOVA cho phép kiểm nghiệm sự khác biệt của các giá trị trung bình trong một giả thuyết chỉ dùng một giá trị α , do đó mức ý nghĩa sẽ nằm ở mức độ kiểm soát được
- Nếu cần kiểm nghiệm theo từng cặp thì mỗi kiểm nghiệm sẽ sử dụng mức ý nghĩa bằng α chia cho số kiểm nghiệm (α/n kiểm nghiệm)
 - Thí dụ nếu quan sát viên cần đánh giá điểm kiểm tra của sinh viên trong lớp theo vị trí trong lớp (bên trái, ở giữa và bên phải) thì sẽ so sánh giá trị trung bình theo từng cặp với mức ý nghĩa là $0.05/3 = 0.017$

ANOVA một chiều

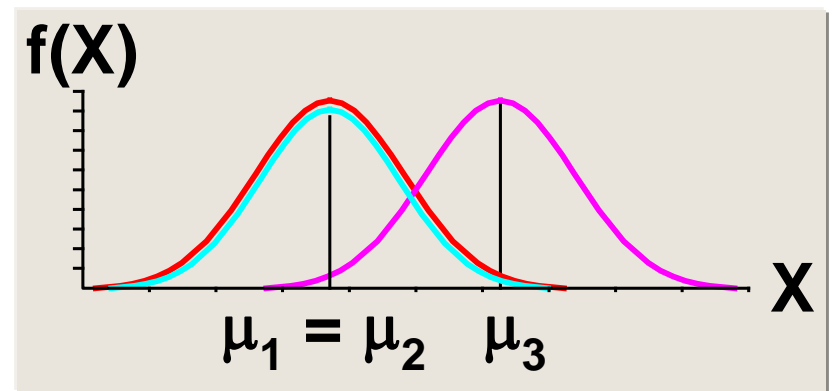
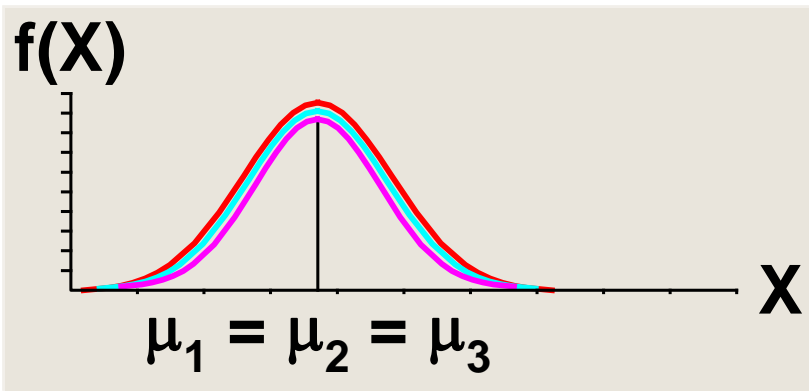
Yếu tố				
1	2	3	4	5
♦	♦	♦	♦	♦
♦	♦	♦	♦	♦
♦	♦	♦	♦	♦
♦	♦	♦	♦	♦
♦	♦	♦	♦	♦
♦	♦	♦	♦	♦
♦		♦	♦	♦
♦		♦	♦	♦
♦		♦	♦	♦
♦		♦		♦
				♦

Các biến trong ANOVA một chiều

- Biến đáp ứng hay biến phụ thuộc là biến mà chúng ta dùng so sánh các nhóm
- Biến yếu tố hay biến độc lập là biến quyết định sử dụng để định nghĩa nhóm (mẫu)
 - Giả sử có k nhóm, thì k là số mức độ của yếu tố
- ANOVA được gọi là một chiều vì các giá trị sắp xếp theo một chiều (chỉ có một biến yếu tố)

Đặt giả thuyết.

- $H_0: \mu_1 = \mu_2 = \mu_3 = \dots = \mu_k$
Thí dụ có 3 nhóm $H_0: \mu_1 = \mu_2 = \mu_3$
- H_1 : có ít nhất một giá trị μ khác với các giá trị khác
Điều này không có nghĩa là $H_1: \mu_1 \neq \mu_2 \neq \mu_3$



Tính các tổng bình phương

- Tổng bình phương chung

$$SST = \sum X_{\text{total}}^2 - (\sum X_{\text{total}})^2 / N$$

- Tổng bình phương giữa các nhóm

$$SSB = \sum [(\sum X_k)^2 / N_k] - (\sum X_{\text{total}})^2 / N$$

- Tổng bình phương trong nhóm. Tính cho từng nhóm và cộng lại

$$SSW_k = \sum X_k^2 - (\sum X_k)^2 / N_k$$

N : tổng số dữ liệu;

N_k : số dữ liệu trong nhóm

k : số nhóm

:

$$N = N_k * k$$

- Ta có: $SST = SSB + SSW$
- Độ tự do
 - Độ tự do của SST là $(N-1)$
 - Độ tự do của SSB là $(k-1)$
 - Độ tự do của SSE là $(N-k)$
- Tính bình phương trung bình
 - $MSB = SSB / (k-1)$
 - $MSE = SSE / (N-k)$
- Tính giá trị Fstat
 - $Fstat = MSB / MSE$
- So sánh Fstat và Ftab. Kết luận

Bảng ANOVA

Nguồn sai số	Tổng bình phương SS	Bậc tự do đf	Bình phương trung bình MS	Giá trị thống kê F
Yếu tố (Between Group)	SSA	k-1	$MSA = \frac{SSA}{k-1}$	$F = \frac{MSA}{MSE}$
Sai số (Within Group)	SSE = SST - SSA	n-k	$MSE = \frac{SSE}{n-k}$	
Tổng cộng	SST	n-1		

Ví dụ:

Hàm lượng Alcaloid (mg) trong một loại dược liệu được thu hái từ 3 vùng khác nhau được số liệu sau:

Vùng 1 : 7,5 6,8 7,1 7,5 6,8 6,6 7,8

Vùng 2 : 5,8 5,6 6,1 6,0 5,7

Vùng 3 : 6,1 6,3 6,5 6,4 6,5 6,3

Hỏi hàm lượng Alcaloid có khác nhau theo vùng hay không?

So sánh 3 loại thuốc bổ A, B, C trên 3 nhóm, người ta được kết quả tăng trọng(kg) như sau:

A: 1,0 1,2 1,4 1,1 0,8 0,6

B: 2,0 1,8 1,9 1,2 1,4 1,0 1,5 1,8

C: 0,4 0,6 0,7 0,2 0,3 0,1 0,2

Hãy so sánh kết quả tăng trọng của 3 loại thuốc bổ trên với $\alpha = 0,01$

Giải:

	Vùng 1	Vùng 2	Vùng 3	
	7,5	5,8	6,1	
	6,8	5,6	6,3	
	7,1	6,1	6,5	
	7,5	6,0	6,4	
	6,8	5,7	6,5	
	6,6		6,3	
	7,8			
n_j	7	5	6	N=18
T_j	50,1	29,2	38,1	T=117,4
$\sum_i x_{ij}^2$	359,79	170,7	242,05	$\sum \sum x_{ij}^2 = 772,54$

$$SST = 772,54 - \frac{(117,4)^2}{18} = 6,831111$$

$$SSA = \frac{(50,1)^2}{7} + \frac{(29,2)^2}{5} + \frac{(38,1)^2}{6} - \frac{(117,4)^2}{18} = 5,326968$$

$$SSE = SST - SSA = 1,5041428$$

Nguồn	SS	Df	MS	F	$F_{k-1; n-k; 1-\alpha}$
Yếu tố	5,326968	2	2,663484	26,561504	3,68
Sai số	1,5041428	15	0,1002761		
Tổng cộng	6,831111	17			

$\Rightarrow F > F_{k-1; n-k; 1-\alpha}$ nên bác bỏ H_0 chấp nhận H_1 .
Vậy hàm lượng Alcaloid có sai khác theo vùng.

Anova: Single Factor

SUMMARY

<i>Groups</i>	<i>Count</i>	<i>Sum</i>	<i>Average</i>	<i>Variance</i>
Vùng 1	7	50.1	7.157143	0.202857
Vùng 2	5	29.2	5.84	0.043
Vùng 3	6	38.1	6.35	0.023

ANOVA

<i>Source of Variation</i>	<i>SS</i>	<i>df</i>	<i>MS</i>	<i>F</i>	<i>P-value</i>	<i>F crit</i>
Between Groups	5.326968	2	2.663484	26.56148	1.17756E-05	3.682316674
Within Groups	1.504143	15	0.100276			
Total	6.831111	17				

3.3. ANOVA hai chiều

- ANOVA hai chiều cho phép khảo sát 2 yếu tố đồng thời, mỗi yếu tố có nhiều mức độ
- ANOVA hai chiều còn cho phép đánh giá được tương tác giữa 2 yếu tố

ANOVA hai chiều

		Yếu tố A			
		1	2	3	4
Yếu tố B	1	♦ ♦ ♦ ♦ ♦	♦ ♦ ♦ ♦ ♦	♦ ♦ ♦ ♦ ♦	♦ ♦ ♦ ♦ ♦
	2	♦ ♦ ♦ ♦ ♦	♦ ♦ ♦ ♦ ♦	♦ ♦ ♦ ♦ ♦	♦ ♦ ♦ ♦ ♦

Phân tích phương sai 2 nhân tố không lặp

Phân tích nhằm đánh giá sự ảnh hưởng của 2 nhân tố (yếu tố) A và B trên các giá trị quan sát x_{ij}

Giả sử nhân tố A có n mức a_1, a_2, \dots, a_n (nhân tố hàng)

B có m mức b_1, b_2, \dots, b_m (nhân tố cột)

* Mẫu điều tra:

A \ B	b_1	b_2	...	b_m
a_1	x_{11}	x_{12}	...	x_{1m}
a_2	x_{21}	x_{22}	...	x_{2m}
:	:	:		:
:	:	:		:
a_n	x_{n1}	x_{n2}	...	x_{nm}

* Giả thiết H_0 :

- Trung bình nhân tố cột bằng nhau
- Trung bình nhân tố hàng bằng nhau
- Không có sự tương tác giữa nhân tố cột và hàng

* Tiên hành tính toán theo bảng dưới đây:

A \ B	b ₁	b ₂	...	b _m	$T_{i^*} = \sum_j x_{ij}$	$\sum_j x_{ij}^2$
a ₁	x ₁₁	x ₁₂	...	x _{1m}	T _{1*}	$\sum_j x_{1j}^2$
a ₂	x ₂₁	x ₂₂	...	x _{2m}	T _{2*}	$\sum_j x_{2j}^2$
:	:	:		:	:	
:	:	:		:	:	
a _n	x _{n1}	x _{n2}	...	x _{nm}	T _{n*}	$\sum_j x_{nj}^2$
$T_{*j} = \sum_i x_{ij}$	T _{*1}	T _{*2}	...	T _{*m}	$T = \sum_{i,j} x_{ij}$	
$\sum_i x_{ij}^2$	$\sum_i x_{i1}^2$	$\sum_i x_{i2}^2$		$\sum_i x_{im}^2$		$\sum_{i,j} x_{ij}^2$

* Bảng ANOVA

Nguồn	SS	df	MS	F
Yếu tố A	$SSA = \frac{\sum_i T_{i*}^2}{m} - \frac{T^2}{m.n}$	n-1	$MS (A) = \frac{SSA}{n-1}$	$F_a = MSA/MSE$
Yếu tố B	$SSB = \frac{\sum_j T_{*j}^2}{n} - \frac{T^2}{m.n}$	m-1	$MSB = \frac{SSB}{m-1}$	$F_b = MSB/MSE$
Sai số	$SSE = SST - SSA - SSB$	(n-1)(m-1)	$MSE = \frac{SSE}{(n-1)(m-1)}$	
Tổng	$SST = \sum_{i,j} x_{ij}^2 - \frac{T^2}{m.n}$	nm-1		

* Kết luận :

- Nếu $F_A > F_{n-1; (n-1)(m-1); 1-\alpha}$ thì bác bỏ yếu tố A (hàng)
- Nếu $F_B > F_{m-1; (n-1)(m-1); 1-\alpha}$ thì bác bỏ yếu tố B (cột)

Ví dụ:

Chiết suất chất X từ 1 loại dược liệu bằng 3 phương pháp và 5 loại dung môi, ta có kết quả:

PP Chiết suất (B) Dung môi (A)	b ₁	b ₂	b ₃
a ₁	120	60	60
a ₂	120	70	50
a ₃	130	60	50
a ₄	150	70	60
a ₅	110	75	54

Hãy xét ảnh hưởng của phương pháp chiết suất và dung môi đến kết quả chiết suất chất X với $\alpha=0,01$

- Giải: Giả thiết H_0 :
- * Trung bình của 3 phương pháp chiết suất bằng nhau
 - * Trung bình của 5 dung môi bằng nhau
 - * Không có sự tương tác giữa phương pháp chiết suất và dung môi

Tính toán:

A \ B	b ₁	b ₂	b ₃	T _{i*}	$\sum_j x_{ij}^2$
a ₁	120	60	60	240	21600
a ₂	120	70	50	240	21800
a ₃	130	60	50	240	23000
a ₄	150	70	60	280	31000
a ₅	110	75	54	239	20641
T* _j	630	335	274	T=1239	
$\sum_i x_{ij}^2$	80300	22625	15116		$\sum_{i,j} x_{ij}^2 = 118041$

$$SST = \sum_{i,j} x_{ij}^2 - \frac{T^2}{m.n} = 118041 - \frac{(1239)^2}{5 \times 3} = 155699,6$$

$$SSA = \frac{\sum_i T_{i^*}^2}{m} - \frac{T^2}{m.n} = \frac{308321}{3} - \frac{(1239)^2}{15} = 432,2667$$

$$SSB = \frac{\sum_j T_{*j}^2}{n} - \frac{T^2}{m.n} = \frac{584201}{5} - \frac{(1239)^2}{15} = 14498,8$$

$$SSE = SST - SSA - SSB = 768,5333$$

Nguồn	SS	df	MS	F
Yếu tố A	SSA= 432,2667	4	MSA = 108,0667	$F_A = 1,1249$
Yếu tố B	SSB= 14498,8	2	MSB = 7249,4	$F_B = 75,4622$
Sai số	SSE= 768,5333	8	MSE = 96,0667	
Tổng	SST = 155699,6	14		

$\Rightarrow F_A < F_{4;8;0,99} = 7,006 \Rightarrow$ Dung môi không ảnh hưởng đến kết quả chiết suất.

$F_B > F_{2;8;0,99} = 8,649 \Rightarrow$ Phương pháp ảnh hưởng đến kết quả chiết suất.

<i>SUMMARY</i>	<i>Count</i>	<i>Sum</i>	<i>Average</i>	<i>Variance</i>
a1	3	240	80	1200
a2	3	240	80	1300
a3	3	240	80	1900
a4	3	280	93.33333333	2433.333333
a5	3	239	79.66666667	800.3333333
b1	5	630	126	230
b2	5	335	67	45
b3	5	274	54.8	25.2

ANOVA

<i>Source of Variation</i>	<i>SS</i>	<i>df</i>	<i>MS</i>	<i>F</i>	<i>P-value</i>	<i>F crit</i>
Rows	432.2666667	4	108.0666667	1.124913255	0.409397603	7.006065061
Columns	14498.8	2	7249.4	75.46217904	6.42093E-06	8.64906724
Error	768.5333333	8	96.06666667			
Total	15699.6	14				

Bài tập

Một nghiên cứu được thực hiện nhằm xem xét sự liên hệ giữa loại phân bón, giống lúa đến năng suất. Năng suất lúa được ghi nhận từ các thực nghiệm sau:

Loại phân bón	Giống lúa	A	B	C
1		65	69	75
2		74	72	70
3		64	68	78
4		83	78	76

Hãy đánh giá sự ảnh hưởng giống lúa, loại phân bón trên năng suất lúa, $\alpha = 0,05$.

Phân tích phương sai 2 nhân tố có lặp

* Mẫu điều tra:

A \ B		B			
		b_1	b_2	...	b_m
a_1		X_{111}	X_{121}		X_{1m1}
		X_{112}	X_{122}		X_{1m2}
		:	:	...	:
		:	:		:
		X_{11r}	X_{12r}		X_{1mr}
a_2		X_{211}	X_{221}	...	X_{2m1}
		X_{212}	X_{222}		X_{2m2}
		:	:		:
		:	:		:
		X_{21r}	X_{22r}		X_{2mr}
:	:	:		:	
:	:	:		:	
a_n		X_{n11}	X_{n21}	...	X_{nm1}
		X_{n12}	X_{n22}		X_{nm2}
		:	:		:
		:	:		:
		X_{n1r}	X_{n2r}		X_{nmr}

* Bảng ANOVA

Nguồn	SS	df	MS	F
Yếu tố A	SSA	n-1	$MSA = \frac{SSA}{n-1}$	$F_A = \frac{MSA}{MSE}$
Yếu tố B	SSB	m-1	$MSB = \frac{SSB}{m-1}$	$F_B = \frac{MSB}{MSE}$
Tương tác AB	SSAB	(n-1)(m-1)	$MSAB = \frac{SSAB}{(n-1)(m-1)}$	$F_{AB} = \frac{MSAB}{MSE}$
Sai số	SSE	nm(r-1)	$MSE = \frac{SSE}{nm(r-1)}$	
Tổng	SST	nmr-1		

* Kết luận

- Nếu $F_A > F_{n-1; nm(r-1); 1-\alpha}$ thì bác bỏ yếu tố A (hàng)
- Nếu $F_B > F_{m-1; nm(r-1); 1-\alpha}$ thì bác bỏ yếu tố B (cột)
- Nếu $F_{AB} > F_{(n-1)(m-1); nm(r-1); 1-\alpha}$ thì có sự tương tác giữa A và B

Ví dụ: Hàm lượng saponin (mg) của cùng một loại dược liệu được thu hái trong 2 mùa (khô và mưa: trong mỗi mùa lấy mẫu 3 lần - đầu mùa, giữa mùa, cuối mùa) và từ 3 miền (Nam, Trung, Bắc) thu được kết quả sau:

Mùa	Thời điểm	Miền		
		Nam	Trung	Bắc
Khô	Đầu mùa	2,4	2,1	3,2
	Giữa mùa	2,4	2,2	3,2
	Cuối mùa	2,5	2,2	3,4
Mưa	Đầu mùa	2,5	2,2	3,4
	Giữa mùa	2,5	2,3	3,5
	Cuối mùa	2,6	2,3	3,5

Hãy cho biết hàm lượng saponin có khác nhau theo mùa hay miền không? Nếu có thì 2 yếu tố mùa và miền có sự tương tác với nhau hay không? $\alpha = 0,05$

Mùa	Miền Nam		Trung		Bắc		$T_{i^{**}}$
Khô	2,4	7,3	2,1	6,5	2,2	9,8	23,6
	2,4		2,2		2,3		
	2,5		2,2		2,3		
Mưa	2,5	7,6	3,2	6,8	3,4	10,4	24,8
	2,5		3,2		3,5		
	2,6		3,4		3,5		
$T_{*j^{*}}$	14,9		13,3		20,2		$T = 48,4$

Tính :

- $\sum_{i,j,k} x_{ijk}^2 = 134,64$
- $\sum_i T_{i^{**}}^2 = 23,6^2 + 24,8^2 = 1172$
- $\sum_j T_{*j^{*}}^2 = 14,9^2 + 13,3^2 + 20,2^2 = 806,94$
- $\sum_{i,j} T_{ij^{*}}^2 = 7,3^2 + 7,6^2 + 6,5^2 + 6,8^2 + 9,8^2 + 10,4^2 = 403,74$
- $T^2 = 48,4^2 = 2342,56$

- $\sum_{i,j,k} x_{ijk}^2 = 134,64$
- $\sum_i T_{i**}^2 = 23,6^2 + 24,8^2 = 1172$
- $\sum_j T_{*j*}^2 = 14,9^2 + 13,3^2 + 20,2^2 = 806,94$
- $\sum_{i,j} T_{ij*}^2 = 7,3^2 + 7,6^2 + 6,5^2 + 6,8^2 + 9,8^2 + 10,4^2 = 403,74$
- $T^2 = 48,4^2 = 2342,56$

$$SST = \sum_{i,j,k} x_{ijk}^2 - \frac{T^2}{nmr} = 134,64 - \frac{2342,56}{18} = 4,4978$$

$$SSA = \frac{\sum_i T_{i**}^2}{mr} - \frac{T^2}{nmr} = \frac{1172}{9} - \frac{2342,56}{18} = 0,08$$

$$SSB = \frac{\sum_j T_{*j*}^2}{nr} - \frac{T^2}{nmr} = \frac{806,94}{6} - \frac{2342,56}{18} = 4,3478$$

$$SSE = \sum_{i,j,k} x_{ijk}^2 - \frac{\sum_{i,j} T_{ij*}^2}{r} = 134,64 - \frac{403,74}{3} = 0,06$$

$$SSAB = SST - SSA - SSB - SSE = 4,4978 - 0,08 - 4,3478 - 0,06 = 0,01$$

Bảng ANOVA

Nguồn	SS	df	MS	F
Yếu tố A (mùa)	0,08	1	0,08	$F_A = 16$
Yếu tố B (miền)	4,3478	2	2,1739	$F_B = 434,78$
Tương tác AB	0,01	2	0,005	$F_{AB} = 1$
Sai số	0,06	12	0,005	
Tổng	4,4978	17		

$\Rightarrow F_A > F_{1; 12; 0,95} = 4,7472$: Hàm lượng saponin khác nhau theo mùa.

$F_B > F_{2; 12; 0,95} = 3,8853$: Hàm lượng saponin khác nhau theo miền.

$F_{AB} < F_{2; 12; 0,95} = 3,8853$: chấp nhận H_0 (không tương tác)

Vậy hàm lượng saponin trong dược liệu khác nhau theo mùa , theo miền và không có sự tương tác giữa mùa và miền trên hàm lượng saponin.

SUMMARY

Nam Trung Bac Total

Count	3	3	3	9
Sum	7.3	6.5	9.8	23.6
Average	2.433333	2.166667	3.266667	2.622222222
Variance	0.003333	0.003333	0.013333	0.251944444

Count	3	3	3	9
Sum	7.6	6.8	10.4	24.8
Average	2.533333	2.266667	3.466667	2.755555556
Variance	0.003333	0.003333	0.003333	0.300277778

Total

Count	6	6	6
Sum	14.9	13.3	20.2
Average	2.483333	2.216667	3.366667
Variance	0.005667	0.005667	0.018667

ANOVA

<i>Source of Variation</i>	<i>SS</i>	<i>df</i>	<i>MS</i>	<i>F</i>	<i>P-value</i>	<i>F crit</i>
Sample	0.08	1	0.08	16	0.001761696	4.747221283
Columns	4.347778	2	2.173889	434.7777778	6.36194E-12	3.885290312
Interaction	0.01	2	0.005	1	0.396569457	3.885290312
Within	0.06	12	0.005			
Total	4.497778	17				

Phân tích phương sai 2 nhân tố có lặp

$$SST = \sum_{i,j,k} (x_{ijk} - \bar{x})^2 = \sum_{i,j,k} x_{ijk}^2 - \frac{T^2}{nmr}$$

$$SSA = mr \sum_i (\bar{x}_{i**} - \bar{x})^2 = \frac{\sum_i T_{i**}^2}{mr} - \frac{T^2}{nmr}$$

$$SSB = nr \sum_j (\bar{x}_{*j*} - \bar{x})^2 = \frac{\sum_j T_{*j*}^2}{nr} - \frac{T^2}{nmr}$$

$$SSAB = r \sum_{j,i} (\bar{x}_{ij*} - \bar{x}_{i**} - \bar{x}_{*j*} + \bar{x})^2 = \frac{\sum_{i,j} T_{ij*}^2}{r} - \frac{\sum_j T_{*j*}^2}{nr} - \frac{\sum_i T_{i**}^2}{mr} + \frac{T^2}{nmr}$$

$$SSE = SST - SSA - SSB - SSAB = \sum_{i,j,k} x_{ijk}^2 - \frac{\sum_{i,j} x_{ij*}^2}{r}$$

* Bảng ANOVA

Nguồn	SS	df	MS	F
Yếu tố A	SSA	n-1	$MSA = \frac{SSA}{n-1}$	$F_A = \frac{MSA}{MSE}$
Yếu tố B	SSB	m-1	$MSB = \frac{SSB}{m-1}$	$F_B = \frac{MSB}{MSE}$
Tương tác AB	SSAB	(n-1)(m-1)	$MSAB = \frac{SSAB}{(n-1)(m-1)}$	$F_{AB} = \frac{MSAB}{MSE}$
Sai số	SSE	nm(r-1)	$MSE = \frac{SSE}{nm(r-1)}$	
Tổng	SST	nmr-1		

* Kết luận

- Nếu $F_A > F_{n-1; nm(r-1); 1-\alpha}$ thì bác bỏ yếu tố A (hàng)
- Nếu $F_B > F_{m-1; nm(r-1); 1-\alpha}$ thì bác bỏ yếu tố B (cột)
- Nếu $F_{AB} > F_{(n-1)(m-1); nm(r-1); 1-\alpha}$ thì có sự tương tác giữa A và B

Bài tập

1) Một nghiên cứu được thực hiện nhằm xem xét sự liên hệ giữa loại phân bón, giống lúa và năng suất. Năng suất lúa được ghi nhận từ các thực nghiệm sau:

Loại phân bón \ Giống lúa	A	B	C
1	65	69	75
	68	71	75
	62	67	78
2	74	72	70
	79	69	69
	76	69	65
3	64	68	78
	72	73	82
	65	75	80
4	83	78	76
	82	78	77
	84	75	75

Hãy cho biết sự ảnh hưởng của loại phân bón, giống lúa trên năng suất, $\alpha = 0,01$

3.4. Qui hoạch hình vuông Latin

- Qui hoạch yếu tố hình vuông Latin là qui hoạch hình vuông trong đó mỗi phần tử được sắp xếp để chỉ xuất hiện 1 lần theo cột hoặc theo hàng
thí dụ hình vuông Latin 3x3

A	B	C
B	C	A
C	A	B

- Trong qui hoạch hình vuông Latin các yếu tố có cùng số mức độ
- Qui hoạch hình vuông Latin 2×2 có thể dùng khảo sát 3 yếu tố, trong đó ảnh hưởng của các tương tác bị lẫn vào ảnh hưởng của các yếu tố. Nói cách khác khi dùng qui hoạch hình vuông Latin thì phải dự đoán trước là các yếu tố không quan trọng.

- Bảng qui hoạch hình vuông Latin 2x2 khảo sát 3 yếu tố

		Yếu tố B	
		b1	b2
Yếu tố A	a1	c1	c2
	a2	c2	c1

- Bảng qui hoạch hình vuông Latin 3x3 khảo sát 3 yếu tố

		Yếu tố B		
		b1	b2	b3
Yếu tố A	a1	c1	c2	c3
	a2	c2	c3	c1
	a3	c3	c1	c2

- Việc phân tích biến lượng của hình vuông Latin tương đương phân tích biến lượng 2 chiều. Đối với yếu tố thứ 2 (yếu tố C) việc phân tích biến lượng sẽ tính tương tự như yếu tố A hoặc B. Ở đây yếu tố C lần với tương tác AB.

Để đơn giản hơn việc phân tích biến lượng tiến hành theo tuần tự như sau:

- Tính tổng theo hàng (cho A) theo cột (cho B) và cho C.
- Tính tổng bình phương tất cả các dữ liệu: SS1
- Tính tổng bình phương chung cho hàng chia cho số dữ liệu trong một hàng: SS2
- Tính tổng bình phương chung cho cột chia cho số dữ liệu trong một cột: SS3
- Tính tổng bình phương chung cho C chia cho số dữ liệu trong một loạt C: SS4

- Tính tổng bình phương toàn thể: $GTSS = SS5$
- Tổng bình phương cho hàng: $SSA = SS2 - SS5$
- Tổng bình phương cho cột: $SSB = SS3 - SS5$
- Tổng bình phương cho C: $SSC = SS4 - SS5$
- Tổng bình phương chung: $SST = SS1 - SS5$
- Tổng bình phương sai số:

$$SSE = SST - SSA - SSB - SSC$$

- Tính MSA , MSB , MSC và MSE
- Tính giá trị F_A , F_B , F_C
- So sánh với giá trị bảng và kết luận

- Bảng ANOVA của qui hoạch hình vuông Latin

Nguồn biến	Độ tự do	Tổng bình phương	Bình phương trung bình	Giá trị F
A	$n - 1$	$SSA = SS2 - SS5$	$S_A^2 = SSA / (n - 1)$	S_A^2 / S_E^2
B	$n - 1$	$SSB = SS3 - SS5$	$S_B^2 = SSB / (n - 1)$	S_B^2 / S_E^2
C	$n - 1$	$SSC = SS4 - SS5$	$S_C^2 = SSC / (n - 1)$	S_C^2 / S_E^2
Sai số	$(n - 1)(n - 2)$	SSE	$S_E^2 = SSE / [(n - 1)(n - 2)]$	
Tổng	$n^2 - 1$	$SST = SS1 - SS5$		

Table 4-11 Coded Data for the Rocket Propellant Problem

Batches of Raw Material	Operators					$y_{i..}$
	1	2	3	4	5	
1	$A = -1$	$B = -5$	$C = -6$	$D = -1$	$E = -1$	-14
2	$B = -8$	$C = -1$	$D = 5$	$E = 2$	$A = 11$	9
3	$C = -7$	$D = 13$	$E = 1$	$A = 2$	$B = -4$	5
4	$D = 1$	$E = 6$	$A = 1$	$B = -2$	$C = -3$	3
5	$E = -3$	$A = 5$	$B = -5$	$C = 4$	$D = 6$	7
$y_{..k}$	-18	18	-4	5	9	10 = $y_{...}$

$$SS_T = \sum_i \sum_j \sum_k y_{ijk}^2 - \frac{y_{...}^2}{N}$$

$$= 680 - \frac{(10)^2}{25} = 676.00$$

$$SS_{\text{Batches}} = \frac{1}{p} \sum_{i=1}^p y_{i..}^2 - \frac{y_{...}^2}{N}$$

$$= \frac{1}{5} [(-14)^2 + 9^2 + 5^2 + 3^2 + 7^2] - \frac{(10)^2}{25} = 68.00$$

$$SS_{\text{Operators}} = \frac{1}{p} \sum_{k=1}^p y_{..k}^2 - \frac{y_{...}^2}{N}$$

$$= \frac{1}{5} [(-18)^2 + 18^2 + (-4)^2 + 5^2 + 9^2] - \frac{(10)^2}{25} = 150.00$$

The totals for the treatments (Latin letters) are

<u>Latin Letter</u>	<u>Treatment Total</u>
<i>A</i>	$y_{.1.} = 18$
<i>B</i>	$y_{.2.} = -24$
<i>C</i>	$y_{.3.} = -13$
<i>D</i>	$y_{.4.} = 24$
<i>E</i>	$y_{.5.} = 5$

The sum of squares resulting from the formulations is computed from these totals as

$$\begin{aligned}SS_{\text{Formulations}} &= \frac{1}{p} \sum_{j=1}^p y_{.j.}^2 - \frac{y_{...}^2}{N} \\ &= \frac{18^2 + (-24)^2 + (-13)^2 + 24^2 + 5^2}{5} - \frac{(10)^2}{25} = 330.00\end{aligned}$$

The error sum of squares is found by subtraction:

$$\begin{aligned}SS_E &= SS_T - SS_{\text{Batches}} - SS_{\text{Operators}} - SS_{\text{Formulations}} \\ &= 676.00 - 68.00 - 150.00 - 330.00 = 128.00\end{aligned}$$

Table 4-12 Analysis of Variance for the Rocket Propellant Experiment

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Square	F_0
Formulations	330.00	4	82.50	7.73
Batches of raw material	68.00	4	17.00	
Operators	150.00	4	37.50	
Error	128.00	12	10.67	
Total	676.00	24		

Table 4-10 Analysis of Variance for the Latin Square Design

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Square	F_0
Treatments	$SS_{\text{Treatments}} = \frac{1}{p} \sum_{j=1}^p y_{.j}^2 - \frac{y_{...}^2}{N}$	$p - 1$	$\frac{SS_{\text{Treatments}}}{p - 1}$	$F_0 = \frac{MS_{\text{Treatments}}}{MS_E}$
Rows	$SS_{\text{Rows}} = \frac{1}{p} \sum_{i=1}^p y_{i.}^2 - \frac{y_{...}^2}{N}$	$p - 1$	$\frac{SS_{\text{Rows}}}{p - 1}$	
Columns	$SS_{\text{Columns}} = \frac{1}{p} \sum_{k=1}^p y_{.k}^2 - \frac{y_{...}^2}{N}$	$p - 1$	$\frac{SS_{\text{Columns}}}{p - 1}$	
Error	SS_E (by subtraction)	$(p - 2)(p - 1)$	$\frac{SS_E}{(p - 2)(p - 1)}$	
Total	$SS_T = \sum_i \sum_j \sum_k y_{ijk}^2 - \frac{y_{...}^2}{N}$	$p^2 - 1$		

3.5. Qui hoạch khối La tin

- Qui hoạch 3 yếu tố, n mức độ ($n > 2$) được thực hiện qua khối vuông. Ba cạnh của khối vuông biểu thị các yếu tố A, B, và C, các mức độ được biểu thị trên các trục.
- Nếu dùng khối vuông Latin để khảo sát 4 yếu tố thì yếu tố thứ 4 – yếu tố D thì mức độ của yếu tố D sẽ được biểu thị tại các điểm tương ứng trên khối vuông và ta có khối Latin bậc nhất.
- Khối latin bậc nhất có thể biểu thị bằng các mặt phẳng song song với mặt trục qua các bảng hoạch định

Khối Latin bậc nhất 3 x 3 x 3

		B		
		A	0	1
C = 0	0	0	1	2
	1	2	0	1
	2	1	2	0

		B			
		A	0	1	2
C = 1	0	0	2	0	1
	1	1	1	2	0
	2	2	0	1	2

		B			
		A	0	1	2
C = 2	0	0	1	2	0
	1	1	0	1	2
	2	2	2	0	1

No	A	B	C	D	y
1	0	0	0	0	Y_1
2	0	1	0	1	Y_2
3	0	2	0	2	Y_3
4	1	0	0	2	Y_4
5	1	1	0	0	Y_5
6	1	2	0	1	Y_6
7	2	0	0	1	Y_7
8	2	1	0	2	Y_8
9	2	2	0	0	Y_9
10	0	0	1	2	Y_{10}
11	0	1	1	0	Y_{11}
12	0	2	1	1	Y_{12}
13	1	0	1	1	Y_{13}
14	1	1	1	2	Y_{14}

No	A	B	C	D	y
15	1	2	1	0	Y_{15}
16	2	0	1	0	Y_{16}
17	2	1	1	1	Y_{17}
18	2	2	1	2	Y_{18}
19	0	0	2	1	Y_{19}
20	0	1	2	2	Y_{20}
21	0	2	2	0	Y_{21}
22	1	0	2	0	Y_{22}
23	1	1	2	1	Y_{23}
24	1	2	2	2	Y_{24}
25	2	0	2	2	Y_{25}
26	2	1	2	0	Y_{26}
27	2	2	2	1	Y_{27}

Cách phân tích biến lượng tiến hành tuần tự như sau:

- Tính tổng của các yếu tố ở từng mức độ
 - A_i ($i = 0, 1, 2, \dots, n-1$)
 - B_j ($j = 0, 1, 2, \dots, n-1$)
 - C_q ($q = 0, 1, 2, \dots, n-1$)
 - D_l ($l = 0, 1, 2, \dots, n-1$)
- Tính tổng bình phương tất cả các số liệu: SS1
- Tính tổng bình phương chung cho yếu tố A chia cho n^2 : SS2
- Tính tổng bình phương chung cho yếu tố B chia cho n^2 : SS3

- Tính tổng bình phương chung cho yếu tố C chia cho n^2 : $SS4$
- Tính tổng bình phương chung cho yếu tố D chia cho n^2 : $SS5$
- Tính tổng bình phương toàn thể: $GTSS = SS6$
- Tổng bình phương cho yếu tố A: $SSA = SS2 - SS6$
- Tổng bình phương cho yếu tố B: $SSB = SS3 - SS6$
- Tổng bình phương cho yếu tố C: $SSC = SS4 - SS6$
- Tổng bình phương cho yếu tố D: $SSD = SS5 - SS6$
- Tổng bình phương chung: $SST = SS1 - SS6$

- Tổng bình phương sai số:

$$SSE = SST - SSA - SSB - SSC - SSD$$

- Tính MSA, MSB, MSC, MSD và MSE
- Tính giá trị F_A, F_B, F_C
- So sánh với giá trị bảng và kết luận

- Bảng ANOVA của qui hoạch khối Latin

Nguồn biến	Độ tự do	Tổng bình phương	Bình phương trung bình	Giá trị F
A	$n - 1$	$SSA = SS2 - SS6$	$S_A^2 = SSA / (n - 1)$	S_A^2 / S_E^2
B	$n - 1$	$SSB = SS3 - SS5$	$S_B^2 = SSB / (n - 1)$	S_B^2 / S_E^2
C	$n - 1$	$SSC = SS4 - SS6$	$S_C^2 = SSC / (n - 1)$	S_C^2 / S_E^2
D	$n - 1$	$SSD = SS5 - SS6$	$S_D^2 = SSD / (n - 1)$	S_D^2 / S_E^2
Sai số	$n^3 - 4n + 3$	SSE	$S_E^2 = SSE / (n^3 - 4n + 3)$	
Tổng	$n^3 - 1$	$SST = SS1 - SS6$		

Thí dụ:

STT	T (A)	Áp suất (B)	Thời gian (C)	Hiệu suất
1	100	20	10	2
2	200	20	10	6
3	100	60	10	4
4	200	60	10	8
5	100	20	20	10
6	200	20	20	18
7	100	60	20	8
8	200	60	20	12

$$SS1 \text{ (Bình phương đáp ứng)} = 752$$

$$SS2 \text{ (Tổng bình phương A/4)} = (24^2 + 44^2)/4 = 628$$

$$SS3 \text{ (Tổng bình phương B/4)} = (36^2 + 32^2)/4 = 580$$

$$SS4 \text{ (Tổng bình phương C/4)} = (20^2 + 48^2)/4 = 676$$

$$SS5 \text{ (Tổng đáp ứng)}^2/8 = (2+6+4+8+10+18+8+12)^2/8 = 578$$

$$SST = SS1 - SS5 = 752 - 578 = 174 \quad (df = 7)$$

$$SSA = SS2 - SS5 = 628 - 578 = 50 \quad (df=1) \quad MSA = 50 \quad FA = 50/6 = 8,33$$

$$SSB = SS3 - SS5 = 580 - 578 = 2 \quad (df=1) \quad MSB = 2 \quad FB = 2/6 = 0,33$$

$$SSC = SS4 - SS5 = 676 - 578 = 98 \quad (df=1) \quad MSC = 98 \quad FC = 98/6 = 16,33$$

$$SSE = SST - SSA - SSB - SSC = 174 - 50 - 2 - 98 = 24 \quad (df=4) \quad MSE = 6$$

F (bảng) (0.05, 1, 4)=7,7