

PHÂN TÍCH PHƯƠNG SAI (ANOVA - ANALYSIS OF VARIANCE)

1. Phân tích phương sai 1 nhân tố

Giả sử nhân tố A có k mức X_1, X_2, \dots, X_k với X_j có phân phối chuẩn $N(a, \sigma^2)$ có mẫu điều tra

X_1	X_2	---	X_k
x_{11}	x_{12}		x_{1k}
x_{21}	x_{22}		x_{2k}
\vdots	\vdots		\vdots
\vdots	\vdots	...	\vdots
x_{n_1}	\vdots		x_{n_k}
	x_{n_2}		

Với mức ý nghĩa α , hãy kiểm định giả thiết:

$$H_0 : a_1 = a_2 = \dots = a_k$$

H_1 : “Tồn tại $j_1 \neq j_2$ sao cho $a_{j_1} \neq a_{j_2}$ “

• Đặt:

▪ Tổng số quan sát: $n = \sum_{j=1}^k n_j$

▪ Trung bình mẫu nhóm j ($j=1, \dots, k$): $\bar{x}_j = \frac{1}{n_j} \sum_{i=1}^{n_j} x_{ij} = \frac{T_j}{n_j}$ với $T_j = \sum_{i=1}^{n_j} x_{ij}$

▪ Trung bình mẫu chung: $\bar{x} = \frac{1}{n} \sum_{j=1}^k \sum_{i=1}^{n_j} x_{ij} = \frac{T}{n}$ với $T = \sum_{j=1}^k \sum_{i=1}^{n_j} x_{ij} = \sum_{j=1}^k T_j$

▪ Phương sai hiệu chỉnh nhóm j : $S_j^2 = \frac{1}{n_j - 1} \sum_{i=1}^{n_j} (x_{ij} - \bar{x}_j)^2$

▪ $SST = \sum_{j=1}^k \sum_{i=1}^{n_j} (x_{ij} - \bar{x})^2$ Tổng bình phương các độ lệch.

▪ $SSA = \sum_{j=1}^k n_j (\bar{x}_j - \bar{x})^2$ Tổng bình phương độ lệch riêng của các nhóm so với \bar{x}

• Tính SST bằng cách chèn thêm \bar{x}_j và khai triển thì được:

$$\begin{aligned} SST &= \sum_{j=1}^k \sum_{i=1}^{n_j} (x_{ij} - \bar{x}_j + \bar{x}_j - \bar{x})^2 = \sum_{j=1}^k \sum_{i=1}^{n_j} (\bar{x}_j - \bar{x})^2 + \sum_{j=1}^k \sum_{i=1}^{n_j} (x_{ij} - \bar{x}_j)^2 \quad \text{vì} \\ &- 2 \sum_{j=1}^k \sum_{i=1}^{n_j} (x_{ij} - \bar{x}_j)(\bar{x}_j - \bar{x}) = -2 \sum_{j=1}^k (\bar{x}_j - \bar{x}) \sum_{i=1}^{n_j} (x_{ij} - \bar{x}_j) = -2 \sum_{j=1}^k (\bar{x}_j - \bar{x}) (\sum_{i=1}^{n_j} x_{ij} - n_j \bar{x}_j) = 0 \\ SST &= \sum_{j=1}^k n_j (\bar{x}_j - \bar{x})^2 + \sum_{j=1}^k \sum_{i=1}^{n_j} (x_{ij} - \bar{x}_j)^2 = SSA + SSE \quad \text{với} \quad SSE = \sum_{j=1}^k \sum_{i=1}^{n_j} (x_{ij} - \bar{x}_j)^2 \end{aligned}$$

* Tổng thứ nhất $SSA = \sum_{j=1}^k n_j (\bar{x}_j - \bar{x})^2$ đặc trưng sự khác nhau giữa các nhóm.

* Tổng thứ hai $\sum_{j=1}^k \sum_{i=1}^{n_j} (x_{ij} - \bar{x}_j)^2$ đặc trưng sự khác nhau giữa số liệu trong nội bộ nhóm.

$SST = \sum_{j=1}^k \sum_{i=1}^{n_j} x_{ij}^2 - \frac{T^2}{n}$	$SSA = \sum_{j=1}^k \frac{T_j^2}{n_j} - \frac{T^2}{n}$	$SSE = SST - SSA$
	$MSA = \frac{SSA}{k-1}$	$MSE = \frac{SSE}{n-k}$

- Nếu H_0 đúng thì $F = \frac{MSA}{MSE}$ có phân phối Fisher bậc tự do k-1; n-k
- Miền B_α : $F > F_{k-1; n-k; 1-\alpha}$

Bảng ANOVA

Nguồn sai số	Tổng bình phương SS	Bậc tự do df	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?

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	$\sum \sum x_{ij}^2 = 772,54$

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

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

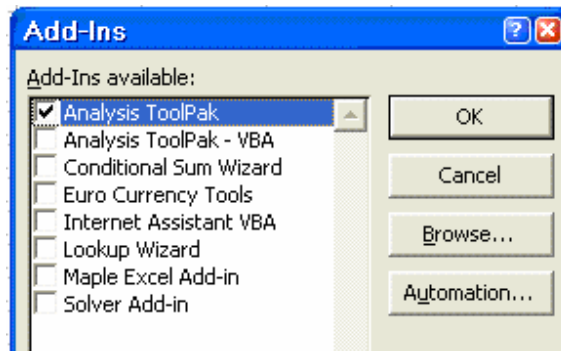
$$SSE = SST - SSA = 1,5041428$$

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			

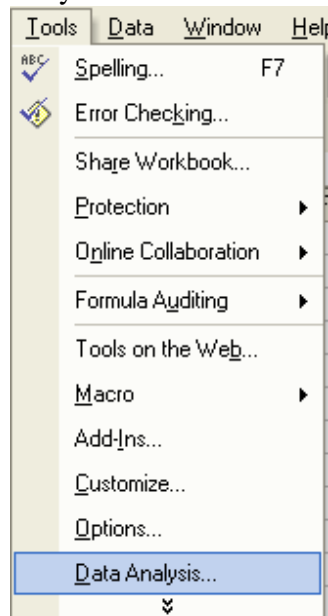
$\Rightarrow F > F_{k-1; n-k; 1-\alpha}$ nên bác bỏ H_0 chấp nhận H_1 .
 Vậy hàm lượng Alcaloid có sai khác theo vùng.

Dùng Excel

1. Nếu trong menu Tools chưa có mục Data Analysis... thì tiến hành cài Analysis ToolPak như sau: Tools \ Add-Ins \ chọn Analysis ToolPak \ OK



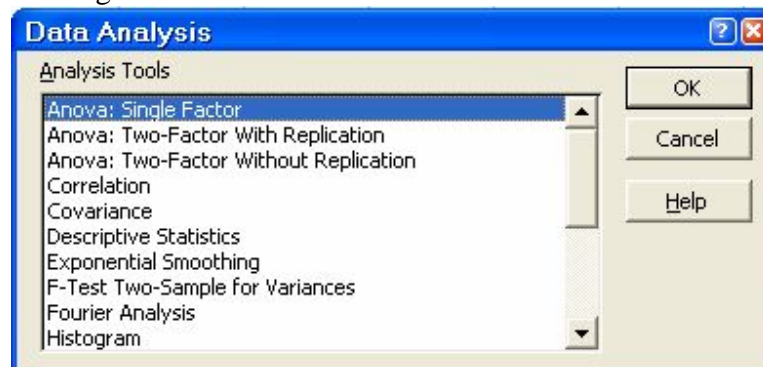
2. Chọn Tools \ Data Analysis ...



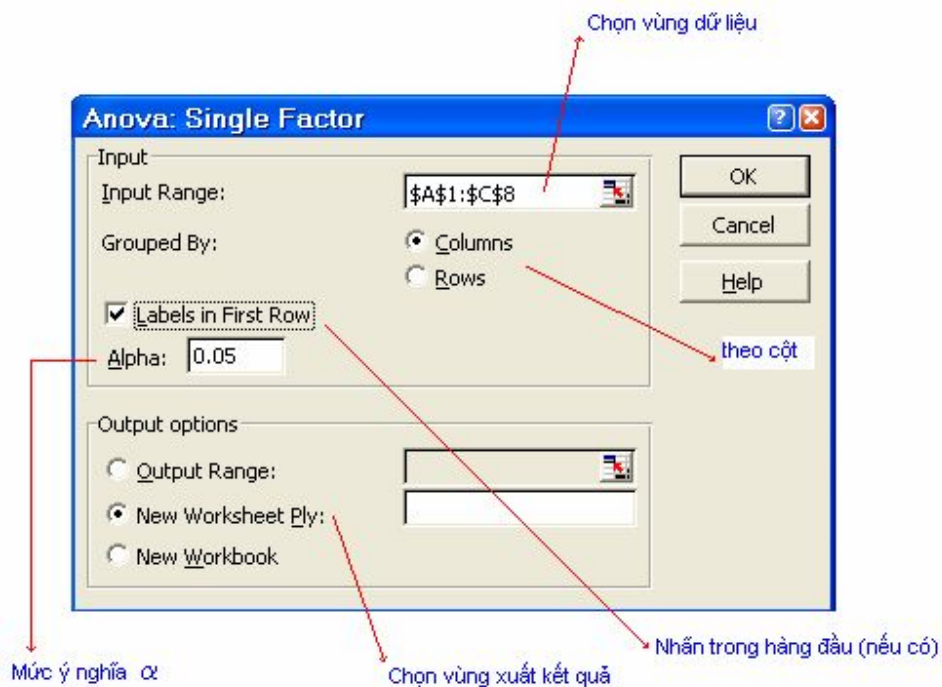
3. Nhập dữ liệu theo cột

	A	B	C
1	Vùng 1	Vùng 2	Vùng 3
2	7.5	5.8	6.1
3	6.8	5.6	6.3
4	7.1	6.1	6.5
5	7.5	6.0	6.4
6	6.8	5.7	6.5
7	6.6		6.3
8	7.8		
9			

4. Chọn mục : Anova: Single Factor



5. Chọn các mục như hình:



6. Kết quả

Anova: Single Factor

SUMMARY

Groups	Count	Sum	Average	Variance
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

Source of Variation	SS	df	MS	F	P-value	F crit
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				

Bài tập

1. 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 $\alpha = 0,01$

2. Một nghiên cứu được thực hiện nhằm xem xét năng suất lúa trung bình của 3 giống lúa. Kết quả thu thập qua 4 năm như sau:

Năm	A	B	C
1	65	69	75
2	74	72	70
3	64	68	78
4	83	78	76

Hãy cho biết năng suất lúa trung bình của 3 giống lúa có khác nhau hay không? $\alpha=0,01$

3. So sánh hiệu quả giảm đau của 4 loại thuốc A, B, C, D bằng cách chia 20 bệnh nhân thành 4 nhóm, mỗi nhóm dùng một loại thuốc giảm đau trên. Kết quả mức độ giảm đau là:

A:	82	89	77	72	92
B:	80	70	72	90	68
C:	77	69	67	65	57
D:	65	75	67	55	63

Hỏi hiệu quả giảm đau của 4 loại thuốc có khác nhau không?

Nếu hiệu quả giảm đau của 4 loại thuốc A, B, C, D khác nhau có ý nghĩa, hãy so sánh từng cặp thuốc với $\alpha = 0,05$

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:

	B	b_1	b_2	...	b_m
A					
a_1		x_{11}	x_{12}	...	x_{1m}
a_2		x_{21}	x_{22}	...	x_{2m}
\vdots		\vdots	\vdots		\vdots
\vdots		\vdots	\vdots		\vdots
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

* Tiến hành tính toán theo bảng dưới đây:

	B	b_1	b_2	...	b_m	$T_{i*} = \sum_j x_{ij}$	$\sum_j x_{ij}^2$
A							
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*}	$\sum_j x_{2j}^2$
\vdots		\vdots	\vdots		\vdots	\vdots	
\vdots		\vdots	\vdots		\vdots	\vdots	
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	$MS_A = \frac{SSA}{n-1}$	$F_A = \frac{SSA}{SSE}$
Yếu tố B	$SSB = \frac{\sum_j T_{*j}^2}{n} - \frac{T^2}{m.n}$	m-1	$MSB = \frac{SSB}{m-1}$	$F_B = \frac{SSB}{SSE}$
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 :

- | |
|---|
| <ul style="list-style-type: none"> • 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 H₀ : * 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		$\sum_{i,j} x_{ij}^2 = 118041$

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

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

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

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

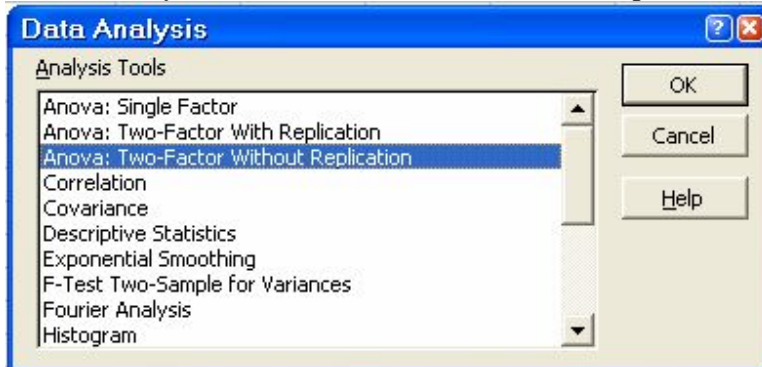
Nguồn	SS	df	MS	F
Yếu tố A	SSA= 432,2667	4	MSA = 108,0667	F _A = 1,1249
Yếu tố B	SSB= 14498,8	2	MSB = 7249,4	F _B = 75,4622
Sai số	SSE= 768,5333	8	MSE = 96,0667	
Tổng	SST = 155699,6	14		

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

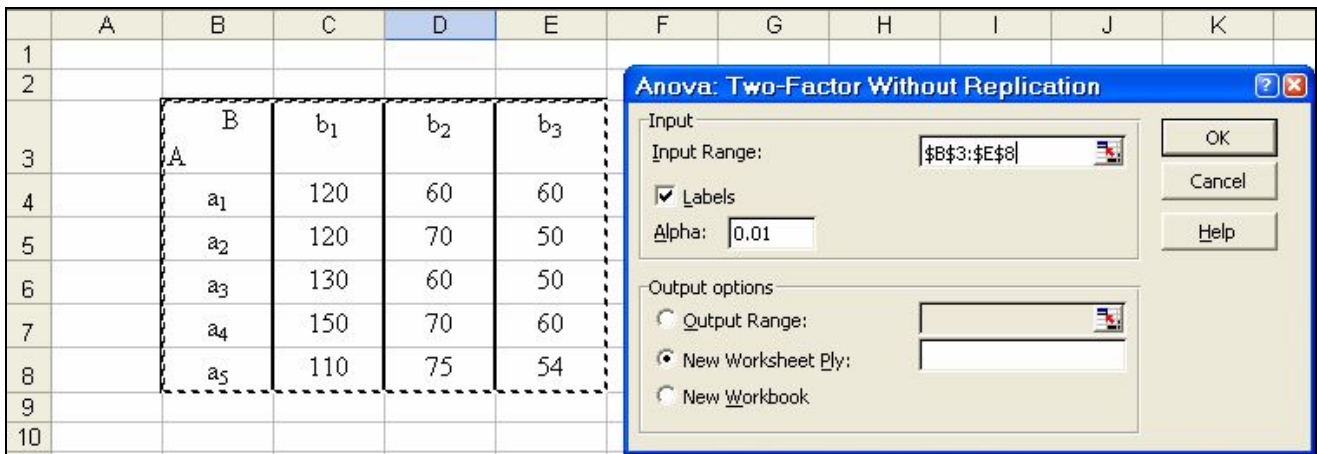
$F_B > F_{2; 8; 0,99} = 8,649 \Rightarrow$ Phương pháp ảnh hưởng đến kết quả chiết suất.

Dùng Excel

- Nhập dữ liệu
- Chọn Tools\Data Analysis...\Anova: Two-Factor without replication



- Chọn các mục như hình



- **Kết quả**
Anova: Two-Factor Without Replication

	<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

- 1) Nghiên cứu về hiệu quả của 3 loại thuốc A, B, C dùng điều trị chứng suy nhược thần kinh. 12 người bệnh được chia làm 4 nhóm theo mức độ bệnh 1, 2, 3, 4; trong mỗi nhóm chia ra để dùng dùng 1 trong 3 loại thuốc trên. Sau 1 tuần điều trị, kết quả đánh giá bằng thang điểm như sau:

Mức độ bệnh \ Thuốc	1	2	3	4
A	25	40	25	30
B	30	25	25	25
C	25	20	20	25

Hãy đánh giá hiệu quả của các loại thuốc A, B, C có khác nhau hay không? với $\alpha = 0,01$

- 2) 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, $\alpha = 0,05$.

- 3) Để khảo sát ảnh hưởng của 4 loại thuốc trừ sâu (1, 2, 3 và 4) và ba loại giống (B1, B2 và B3) đến sản lượng của cam, các nhà nghiên cứu tiến hành một thí nghiệm loại giai thừa. Trong thí nghiệm này, mỗi giống cam có 4 cây cam được chọn một cách ngẫu nhiên, và 4 loại thuốc trừ sâu áp dụng (cũng ngẫu nhiên) cho mỗi cây cam.

Kết quả nghiên cứu (sản lượng cam) cho từng giống và thuốc trừ sâu như sau:

Thuốc trừ sâu \ Giống Cam	1	2	3	4
B1	29	50	43	53
B2	41	58	42	73
B3	66	85	63	85

Hãy cho biết thuốc trừ sâu, giống cam có ảnh hưởng đến sản lượng cam không? $\alpha = 0,05$

- 4) 4 chuyên gia tài chính được yêu cầu dự đoán về tốc độ tăng trưởng (%) trong năm tới của 5 công ty trong ngành nhựa. Dự đoán được ghi nhận như sau:

Công ty	Chuyên gia			
	A	B	C	D
1	8	12	8,5	13
2	14	10	9	11
3	11	9	12	10
4	9	13	10	13
5	12	10	10	10

Hãy lập bảng ANOVA. Có thể nói rằng dự đoán tốc độ tăng trưởng trung bình là như nhau cho cả 5 công ty nhựa được không?

3. Phân tích phương sai 2 nhân tố có lặp

Tương tự như bài toán phân tích phương sai 2 nhân tố không lặp, chỉ khác mỗi mức $((a_i, b_j))$ đều có sự lặp lại r lần thí nghiệm và ta cần khảo sát thêm sự tương tác (interaction term) F_{AB} giữa 2 nhân tố A và B.

* 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}
		\vdots	\vdots	...	\vdots
		\vdots	\vdots		\vdots
		X_{11r}	X_{12r}		X_{1mr}
a_2		X_{211}	X_{221}	...	X_{2m1}
		X_{212}	X_{222}		X_{2m2}
		\vdots	\vdots		\vdots
		\vdots	\vdots		\vdots
		X_{21r}	X_{22r}		X_{2mr}
\vdots	\vdots	\vdots		\vdots	
\vdots	\vdots	\vdots		\vdots	
a_n		X_{n11}	X_{n21}	...	X_{nm1}
		X_{n12}	X_{n22}		X_{nm2}
		\vdots	\vdots		\vdots
		\vdots	\vdots		\vdots
		X_{n1r}	X_{n2r}		X_{nmr}

* Xử lý mẫu: Tính tổng hàng $T_{i**} = \sum_{j,k} x_{ijk}$, tổng cột $T_{*j*} = \sum_{i,k} x_{ijk}$

A \ B		B				T_{i**}
		b_1	b_2	...	b_m	
a_1		X_{111}	X_{121}		X_{1m1}	$T_{1**} = \sum_{j,k} x_{1jk}$
		X_{112}	X_{122}		X_{1m2}	
		\vdots	\vdots	...	\vdots	
		\vdots	\vdots		\vdots	
		X_{11r}	X_{12r}		X_{1mr}	
a_2		X_{211}	X_{221}	...	X_{2m1}	$T_{2**} = \sum_{j,k} x_{2jk}$
		X_{212}	X_{222}		X_{2m2}	
		\vdots	\vdots		\vdots	
		\vdots	\vdots		\vdots	
		X_{21r}	X_{22r}		X_{2mr}	
\vdots	\vdots	\vdots		\vdots		
\vdots	\vdots	\vdots		\vdots		
a_n		X_{n11}	X_{n21}	...	X_{nm1}	$T_{n**} = \sum_{j,k} x_{njk}$
		X_{n12}	X_{n22}		X_{nm2}	
		\vdots	\vdots		\vdots	
		\vdots	\vdots		\vdots	
		X_{n1r}	X_{n2r}		X_{nmr}	
T_{*j*}	$T_{*1*} = \sum_{i,k} x_{i1k}$	$T_{*2*} = \sum_{i,k} x_{i2k}$		$T_{*m*} = \sum_{i,k} x_{imk}$	$T = \sum_{i,j,k} x_{ijk}$	

Cần tính: $\sum_{i,j,k} x_{ijk}^2$ $\sum_i T_{i**}^2$ $\sum_j T_{*j*}^2$ $\sum_{i,j} T_{ij*}^2$

Suy ra

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

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

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

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

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

* 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

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? $\alpha = 0,05$

Giải:

Mùa	Miền		Nam		Trung		Bắc		$T_{i^{**}}$
Khô	2,4	7,3	2,1	6,5	2,2	9,8	2,3	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	3,5	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 :

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

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

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

$$SSB = \frac{\sum_j T_{*j^{**}}^2}{nr} - \frac{T^2}{nmr} = \frac{826,94}{6} - \frac{2342,56}{18} = 4,3478$$

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

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

Bảng ANOVA

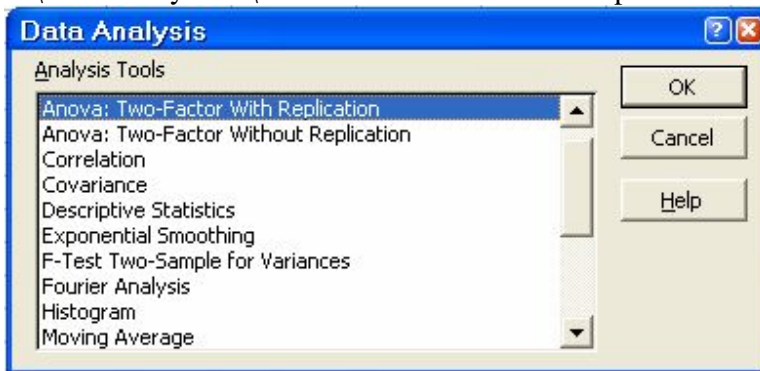
Nguồn	SS	df	MS	F
Yếu tố A (mùa)	0,08	1	0,08	$F_A = 16$
Yếu tố B (miền)	4,3478	2	2,1739	$F_B = 434,78$
Tương tác AB	0,01	2	0,005	$F_{AB} = 1$
Sai số	0,06	12	0,005	
Tổng	4,4978	17		

- $\Rightarrow F_A > F_{1; 12; 0,95} = 4,7472$: 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.

Dùng EXCEL

- * Nhập dữ liệu
- * Chọn Tools\Data Analysis...\Anova: Two Factor With Replication



- * Chọn các mục như hình

* Bảng ANOVA
Anova: Two-Factor With Replication

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

<i>Total</i>				
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						
Source of Variation	SS	df	MS	F	P-value	F crit
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				

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	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, $\alpha = 0,01$

- 2) Điều tra mức tăng trưởng chiều cao của 1 loại cây trồng theo loại đất trồng và loại phân bón có kết quả:

Loại phân	Loại đất	1	2	3
	A		5,5	4,5
		5,5	4,5	4,0
		6,0	4,0	3,0
B		5,6	5,0	4,0
		7,0	5,5	5,0
		7,0	5,0	4,5

Hỏi có sự khác nhau của mức tăng trưởng chiều cao theo loại đất và loại phân bón ? $\alpha=0,05$

- 3) Nghiên cứu sản lượng bông (tạ/ha) theo mật độ trồng A và phân bón B thu được:

Mật độ trồng	Phân bón			
	b1	b2	b3	b4
a1	16	19	19	20
	14	20	21	24
	21	23	22	21
	16	19	20	17
a2	17	19	21	20
	15	18	21	20
	17	18	22	22
	19	20	23	19
a3	18	20	22	25
	18	23	18	22
	19	21	21	21
	17	21	21	23

Hỏi có sự khác nhau của sản lượng bông theo mật độ trồng, theo phân bón với mức $\alpha=0,05$

BÀI TẬP

- 1) Một nhà máy thủy điện sử dụng các turbines được giải nhiệt bằng nước. Nếu nước được dung để giải nhiệt bị ô nhiễm thì hệ thống máy móc sẽ bị xói mòn. Do đó, người ta sử dụng các máy lọc để làm giảm mức ô nhiễm của nước. Giám đốc nhà máy muốn trắc nghiệm tính hiệu quả của 4 máy lọc đang sử dụng. Ở mỗi máy lọc người ta lấy ngẫu nhiên độc lập nhau 3 mẫu nước đã được lọc và đo mức độ ô nhiễm. Các kết quả có được như sau:

Máy lọc 1	Máy lọc 2	Máy lọc 3	Máy lọc 4
10	11	13	23
9	16	8	18
5	9	9	25

- 2) Một nghiên cứu được thực hiện để so sánh tuổi thọ (giờ) của 4 nhãn hiệu Pin: A, B, C, D. Kết quả ghi nhận được như sau:

Hiệu A	Hiệu B	Hiệu C	Hiệu D
15	14	19	16
16	15	20	15
18	16	16	16
20	15	13	18
19	14	17	
20			

Yêu cầu: Giả định tuổi thọ pin có phân phối chuẩn, phương sai bằng nhau. Với phương pháp ANOVA, ở mức ý nghĩa 0,05, có thể kết luận rằng tuổi thọ trung bình của 4 nhãn hiệu pin là không khác nhau được không?

- 3) Ba mẫu thiết kế bao bì của một loại sản phẩm được xem xét bằng cách thu thập doanh số (triệu đồng/tuần) của mỗi loại bao bì trong một mẫu ngẫu nhiên các cửa hàng. Kết quả được ghi nhận trong bảng sau:

Mẫu bao bì I	Mẫu bao bì II	Mẫu bao bì III
18	24	19
16	25	24
29	21	24
26	31	28
29	22	15
14		29
12		32
23		

Với kiểm định ANOVA ở mức ý nghĩa 0,01, có thể kết luận rằng các mẫu bao bì không ảnh hưởng đến doanh số được không? (Giả định doanh số theo các mẫu bao bì có phân phối chuẩn, phương sai bằng nhau).

- 4) Một nhà sản xuất muốn kiểm tra xem 3 máy có công suất khác nhau không. Ông ta chỉ định ngẫu nhiên 15 công nhân được đào tạo cùng một phương pháp làm việc trên 3 máy (5 người/máy). Với mức rủi ro 5%, liệu 3 máy có công suất khác nhau?

Máy 1	Máy 2	Máy 3
25.40	23.40	20.00
26.31	21.80	22.20
24.10	23.50	19.75
23.74	22.75	20.60
25.10	21.60	20.40

- 5) Để so sánh hiệu năng của 3 loại thuốc diệt muỗi A, B, C người ta thực hiện một thực nghiệm như sau: Có 21 thùng, mỗi thùng nhốt vài trăm con muỗi. Chia ngẫu nhiên các thùng này thành 3 nhóm, mỗi nhóm 7 thùng. Muỗi ở trong mỗi nhóm thùng được xịt một loại thuốc khác nhau A, B hoặc C, tỉ lệ % muỗi chết được ghi nhận như sau:

Thuốc diệt muỗi A	Thuốc diệt muỗi B	Thuốc diệt muỗi C
68	58	71
80	60	62
69	70	58
76	51	74
68	57	65
77	71	59
60	61	57

Với kiểm định ANOVA ở mức ý nghĩa $\alpha = 0,05$, có thể nói khả năng diệt muỗi (thể hiện thông qua tỉ lệ muỗi chết trung bình) của 3 loại thuốc là như nhau được không? (giả định muỗi chết có phân phối chuẩn, phương sai bằng nhau).

- 6) Trưởng phòng kỹ thuật của một nhà máy sản xuất vỏ xe thực hiện một nghiên cứu để đánh giá sự khác biệt về chất lượng sản phẩm giữa 3 ca sản xuất: sáng, chiều, đêm. Chọn ngẫu nhiên một số sản phẩm để kiểm tra, kết quả ghi nhận như sau:

Thời gian sản xuất	Số sản phẩm	Độ bền trung bình (ngàn km)	Tổng bình phương các sai lệch
Sáng	10	25,95	6,255
Chiều	12	25,50	6,595
Tối	15	23,75	7,555

Yêu cầu: Với mức ý nghĩa tùy theo quyết định của Anh (chị), có thể kết luận rằng có sự khác biệt về độ bền giữa các sản phẩm sản xuất ra ở ca sáng, ca chiều và ca đêm hay không? Nếu có, sự khác biệt đó như thế nào?

- 7) Bốn trạm sửa chữa và bảo hành xe Honda trong một thành phố lớn tuyên bố rằng khách hàng sẽ được phục vụ nhanh chóng ngay khi xe được đưa tới trạm. Giám đốc phụ trách dịch vụ hậu mãi của hãng tiến hành kiểm tra chất lượng dịch vụ của các trạm bảo hành, bằng cách chọn ngẫu nhiên khách hàng đến trạm trong giờ cao điểm (9 đến 11 giờ sáng) và ghi nhận thời gian chờ đợi của họ. Một phần kết quả tính toán cho trong bảng sau:

Trạm bảo hành	Số khách hàng	Thời gian chờ TB (phút)	Phương sai
A	3	5,133333	0,323333
B	4	8	1,433333
C	5	5,04	0,748
D	4	6,475	0,595833

Lập bảng ANOVA. Số liệu trên có chứng tỏ rằng thời gian chờ đợi của khách hàng ở các trạm bảo hành của hãng là không khác nhau? Kết luận với mức ý nghĩa 0,05.

- 8) Một hãng sản xuất ô tô thực hiện một nghiên cứu để đo lường sự khác biệt mức nhiên liệu tiêu thụ trung bình giữa 3 loại xe: cỡ nhỏ (4 chỗ), trung bình (8 chỗ), và xe cỡ lớn (12 chỗ). Chọn ngẫu nhiên 27 xe, kết quả tính toán cho trong bảng sau:

Loại xe	Số xe	Mức nhiên liệu tiêu thụ TB (lit/100km)	Phương sai
Nhỏ	12	8,133333	2,343333
Trung Bình	9	9,583253	2,453333
Lớn	6	10,04578	3,74853

Lập bảng ANOVA. Số liệu trên có chứng tỏ rằng mức nhiên liệu tiêu thụ trung bình của các loại xe là không khác nhau? Kết luận với mức ý nghĩa 0,05.

- 9) Một phần bảng ANOVA về ảnh hưởng của loại phân bón với các giống lúa khác nhau đến năng suất lúa được trình bày sau đây:

Biến thiên	Bậc tự do	Tổng bình phương các sai lệch	Trung bình các độ lệch bình phương (phương sai)	Giá trị F
Giữa các nhóm	5	605		
Giữa các khối	2	245		
Sai số		150		
Tổng				

Yêu cầu: Hãy xác định:

Tổng số quan sát khi thực hiện cuộc nghiên cứu trên.

Hoàn tất bảng ANOVA.

Đặt giả thuyết H_0 và H_1 .

Kiểm định các giả thuyết với mức ý nghĩa $\alpha = 5\%$.

10) Kết quả tính toán cho trong bảng ANOVA như sau:

Sources	Df	Sum of Squares	Mean Squares	F
Between Groups	4	501	1225,25	9,109
Between Blocks	2	225	112,50	8,182
Error	8	110	13,75	
Total	14	836		

Yêu cầu: Hãy xác định

Tổng số quan sát khi thực hiện nghiên cứu trên.

Phát biểu giả thuyết.

Sử dụng mức ý nghĩa 0,01, hãy kết luận trung bình của các tổng thể.

11) 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:

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

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:

12) Để khảo sát ảnh hưởng của 4 loại thuốc trừ sâu (1, 2, 3 và 4) và ba loại giống (B1, B2 và B3) đến sản lượng của cam, các nhà nghiên cứu tiến hành một thí nghiệm loại giai thừa. Trong thí nghiệm này, mỗi giống cam có 4 cây cam được chọn một cách ngẫu nhiên, và 4 loại thuốc trừ sâu áp dụng (cũng ngẫu nhiên) cho mỗi cây cam.

Kết quả nghiên cứu (sản lượng cam) cho từng giống và thuốc trừ sâu như sau:

Thuốc trừ sâu Giống Cam	1	2	3	4
B1	29	50	43	53
B2	41	58	42	73
B3	66	85	63	85

Hãy cho biết thuốc trừ sâu, giống cam có ảnh hưởng đến sản lượng cam không? $\alpha = 0,05$

- 13) 4 chuyên gia tài chính được yêu cầu dự đoán về tốc độ tăng trưởng (%) trong năm tới của 5 công ty trong ngành nhựa. Dự đoán được ghi nhận như sau:

Công ty	Chuyên gia			
	A	B	C	D
1	8	12	8,5	13
2	14	10	9	11
3	11	9	12	10
4	9	13	10	13
5	12	10	10	10

Hãy lập bảng ANOVA. Có thể nói rằng dự đoán tốc độ tăng trưởng trung bình là như nhau cho cả 5 công ty nhựa được không?

- 14) Một công ty vận chuyển thực hiện một nghiên cứu để xem xét ảnh hưởng của lộ trình đến thời gian vận chuyển (phút) giữa 2 địa điểm. Số liệu thống kê về thời gian vận chuyển của 9 chuyến trong một tuần được thực hiện trên các lộ trình và thời gian khác nhau trong ngày cho trong bảng sau:

Thời gian	Lộ trình		
	A	B	C
10 - 12 giờ sáng	50	52	54
1 - 3 giờ chiều	45	65	62
7 -10 giờ tối	55	47	50

Yêu cầu: Ở mức ý nghĩa 5%, hãy kết luận xem:

Có sự khác biệt về thời gian vận chuyển trung bình giữa 3 lộ trình hay không? Nếu có, công ty nên chọn lộ trình nào?

Có sự khác biệt về thời gian vận chuyển trung bình giữa các thời gian khác nhau trong ngày hay không? Nếu có, công ty nên thực hiện vận chuyển vào thời gian nào?