

## *Chương 3*

---

---

**ÁP DỤNG MS-EXCEL**

**TRONG QUY HOẠCH THỰC NGHIỆM VÀ PHÂN TÍCH PHƯƠNG SAI**

- ☐ **Phân tích phương sai một yếu tố**
- ☐ **Phân tích phương sai hai yếu tố**
  - **Không lặp**
  - **Có lặp**
- ☐ **Phân tích phương sai ba yếu tố**

## A- PHÂN TÍCH PHƯƠNG SAI

### MỘT YẾU TỐ

#### 5.1 Khái niệm thống kê

Phép phân tích phương sai được dùng trong các trắc nghiệm để so sánh các giá trị trung bình của *hai* hay *nhiều* mẫu được lấy từ các phân số. Đây có thể được xem như phần mở rộng của trắc nghiệm t hay z (*so sánh hai giá trị trung bình*).

Mục đích của sự phân tích phương sai một yếu tố là đánh giá sự ảnh hưởng của một yếu tố (nhân tạo hay tự nhiên) nào đó trên các giá trị quan sát,  $Y_i$ , ( $i = 1, 2, \dots, k$ ).

#### Mô hình

	Yếu tố thí nghiệm				
	1	2	...	k	
	$Y_{11}$	$Y_{21}$	...	$Y_{k1}$	
	$Y_{12}$	$Y_{22}$	...	$Y_{k2}$	
	...	...	...	...	
	$Y_{1N}$	$Y_{2N}$	...	$Y_{kN}$	
<b>Tổng cộng</b>	$T_1$	$T_2$	...	$T_k$	$T$
<b>Trung bình</b>	$\bar{Y}_1$	$\bar{Y}_2$	...	$\bar{Y}_k$	$\bar{Y}$

#### Bảng ANOVA

Nguồn sai số	Bậc tự do	Tổng số bình phương	Bình phương trung bình	Giá trị thống kê
<b>Yếu tố</b>	$k-1$	$SSF = \sum_{i=1}^k \frac{T_i^2}{N} - \frac{T^2}{N}$	$MSF = \frac{SSF}{k-1}$	$F = \frac{MSF}{MSE}$
<b>Sai số</b>	$N-k$	$SSE = SST - SSF$	$MSE = \frac{SSE}{N-k}$	
<b>Tổng cộng</b>	$N-1$	$SST = \sum_{i=1}^k \sum_{j=1}^N Y_{ij}^2 - \frac{T^2}{N}$		

#### Trắc nghiệm

\* Giả thuyết:

$H_0: \mu_1 = \mu_2 = \dots \mu_k \Leftrightarrow$  “Các giá trị trung bình bằng nhau”

$H_1: \mu_i \neq \mu_j \Leftrightarrow$  “Ít nhất có hai giá trị trung bình khác nhau”

\* Giá trị thống kê:  $F = \frac{MSF}{MSE}$

\* Biện luận:

Nếu  $F < F_{\alpha}(k-1; N-k) \Rightarrow$  Chấp nhận giả thuyết  $H_0$ .

## 5.2 Áp dụng MS-EXCEL

**Thí dụ 12:** Hàm lượng alcaloid (mg) trong một loại dược liệu được thu hái từ ba vùng khác nhau được trình bày trong bảng sau:

Vùng I	Vùng II	Vùng III
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		

Hàm lượng alcaloid có khác nhau theo vùng?

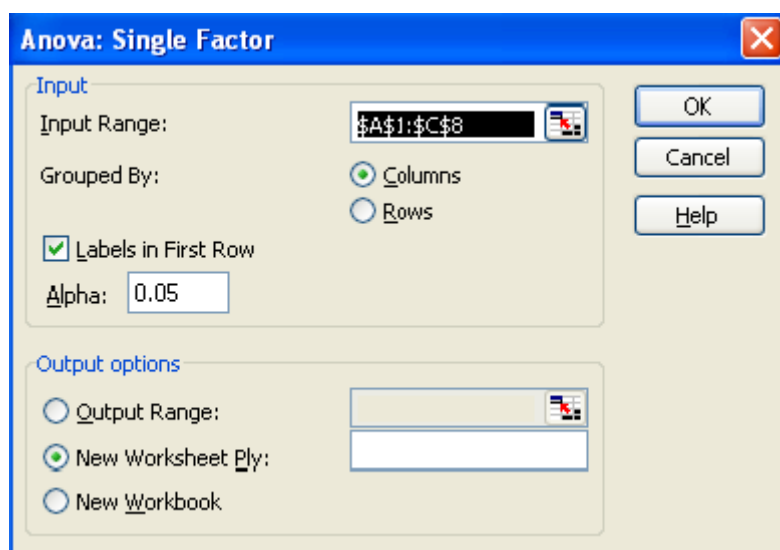
**Nhập dữ liệu vào bảng tính**

	A	B	C
1	Vùng I	Vùng II	Vùng III
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		

**Áp dụng “Anova: Single Factor”**

- Nhập lần lượt đơn lệnh *Tools* và lệnh *Data Analysis*.
- Chọn chương trình *Anova: Single Factor* trong hộp thoại *Data Analysis* rồi nhấp nút *OK*.
- Trong hộp thoại *Anova: Single Factor*, lần lượt ấn định:
  - Phạm vi đầu vào (*Input Range*)
  - Cách sắp xếp theo hàng hay cột (*Group By*)
  - Nhãn dữ liệu (*Labels in Fisrt Row/Column*)

Anova: Single Factor						
SUMMARY						
Groups	Count	Sum	Average	Variance		
Vùng I	7	50.1	7.157143	0.202857		
Vùng II	5	29.2	5.84	0.043		
Vùng III	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	1.18E-05	3.68232
Within Groups	1.504143	15	0.100276			
Total	6.831111	17				



Hình 5.1 Hộp thoại *Anova: Single Factor*

### Kết quả và biện luận

$$F = 26.561 > F_{0.05} = 3.682$$

⇒ Bác bỏ giả thuyết  $H_0$ .

Vậy hàm lượng alkaloid khác nhau theo vùng.

## B- PHÂN TÍCH PHƯƠNG SAI HAI YẾU TỐ (KHÔNG LẬP)

### 5.3 Khái niệm thống kê

Sự phân tích này nhằm đánh giá sự ảnh hưởng của hai yếu tố trên các giá trị quan sát  $Y_{ij}$  ( $i = 1, 2, \dots, r$ : yếu tố A;  $j = 1, 2, \dots, c$ : yếu tố B).

*Mô hình*

Yếu tố A	Yếu tố B				Tổng cộng	Trung bình
	1	2	...	c		
1	$Y_{11}$	$Y_{12}$	...	$Y_{1c}$	$Y_{1.}$	$\bar{Y}_1$
2	$Y_{21}$	$Y_{22}$	...	$Y_{2c}$	$Y_{2.}$	$\bar{Y}_2$
...	...	...	...	...	...	...
r	$Y_{r1}$	$Y_{r2}$	...	$Y_{rc}$	$Y_{r.}$	$\bar{Y}_r$
Tổng cộng	$T_{.1}$	$T_{.2}$	...	$T_{.c}$	$T_{..}$	
Trung bình	$\bar{Y}_{.1}$	$\bar{Y}_{.2}$	...	$\bar{Y}_{.c}$	$\bar{Y}_{..}$	

**Bảng ANOVA**

Nguồn sai số	Bậc tự do	Tổng số bình phương	Bình phương trung bình	Giá trị thống kê
<b>Yếu tố A (Hàng)</b>	(r-1)	$SSB = \sum_{i=1}^r \frac{T_i^2}{c} - \frac{T_{..}^2}{rc}$	$MSB = \frac{SSB}{(r-1)}$	$F_R = \frac{MSB}{MSE}$
<b>Yếu tố B (Cột)</b>	(c-1)	$SSB = \sum_{j=1}^c \frac{T_j^2}{r} - \frac{T_{..}^2}{rc}$	$MSF = \frac{SSF}{(c-1)}$	$F_C = \frac{MSF}{MSE}$
<b>Sai số</b>	(r-1)(c-1)	$SSE = SST - (SSF + SSB)$	$MSB = \frac{SSB}{(r-1)}$	
<b>Tổng cộng</b>	(rc-1)	$SST = \sum_{i=1}^r \sum_{j=1}^c Y_{ij}^2 - \frac{T_{..}^2}{r}$		

### Trắc nghiệm

\* Giả thuyết:

$H_0: \mu_1 = \mu_2 = \dots \mu_k \Leftrightarrow$  “Các giá trị trung bình bằng nhau”

$H_1: \mu_i \neq \mu_j \Leftrightarrow$  “Ít nhất có hai giá trị trung bình khác nhau”

\* Giá trị thống kê:  $F_R = \frac{MSB}{MSE}$  và  $F_C = \frac{MSF}{MSE}$

\* Biện luận:

Nếu  $F_R < F_a[b-1, (k-1)(b-1)] \Rightarrow$  Chấp nhận  $H_0$  (yếu tố A)

Nếu  $F_C < F_a[k-1, (k-1)(b-1)] \Rightarrow$  Chấp nhận  $H_0$  (yếu tố B).

#### 5.4 Áp dụng MS-EXCEL

**Thí dụ 13:** Hàm lượng flavonoid (mg) trong cùng một mẫu dược liệu được chiết xuất bởi 5 phương pháp với 5 loại dung môi khác nhau:

Phương pháp	Dung môi				
	I	II	III	IV	V
A	12.9	17.1	11.6	23.4	17.6
B	13.4	18.1	19.6	22.1	16.8
C	15.6	16.9	16.8	21.5	18.1
D	12.7	17.8	21.3	20.9	17.9
E	13.5	19.3	18.7	23.2	17.3

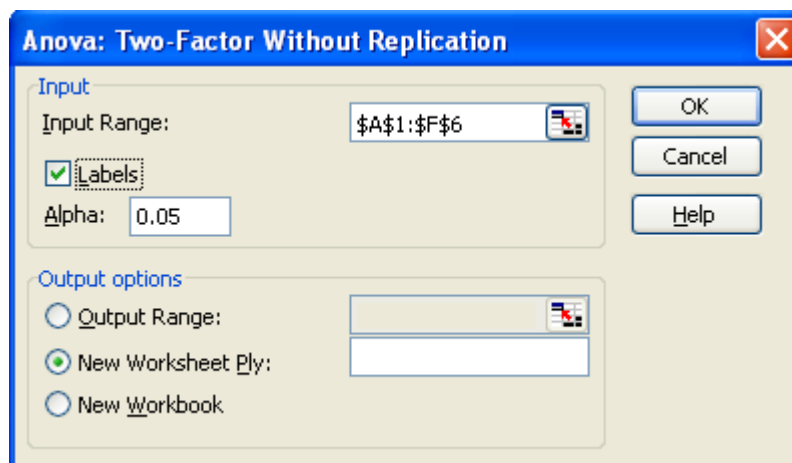
Phương pháp hay dung môi có ảnh hưởng đến kết quả của sự chiết xuất dược liệu nêu trên?

**Nhập dữ liệu vào bảng tính**

	A	B	C	D	E	F
1	I	II	III	IV	V	
2	A	12.9	17.1	11.6	23.4	17.6
3	B	13.4	18.1	19.6	22.1	16.8
4	C	15.6	16.9	16.8	21.5	18.1
5	D	12.7	17.8	21.3	20.9	17.9
6	E	13.5	19.3	18.7	23.2	17.3

*Áp dụng “Anova: Two-Factor Without Replication”*

- Nhấp lần lượt đơn lệnh *Tools* và lệnh *Data Analysis*.
- Chọn chương trình *Anova: Two-* trong hộp thoại *Data Analysis* rồi nhấp nút *OK*.
- Trong hộp thoại *Anova: Two-Factor Without Replication*. lần lượt ấn định các chi tiết:
  - Phạm vi đầu vào (*Input Range*).
  - Nhãn dữ liệu (*Labels in Fisrt Row/Column*)
  - Ngưỡng tin cậy (*Alpha*).
  - Phạm vi đầu ra (*Output Range*).



Hình 5.2: Hộp thoại Anova: Two-Factor Without Replication

Anova: Two-Factor Without Replication						
SUMMARY	Count	Sum	Average	Variance		
A	5	82.6	16.52	21.537		
B	5	90	18	10.495		
C	5	88.9	17.78	5.107		
D	5	90.6	18.12	11.842		
E	5	92	18.4	12.29		
I	5	68.1	13.62	1.337		
II	5	89.2	17.84	0.908		
III	5	88	17.6	13.885		
IV	5	111.1	22.22	1.157		
V	5	87.7	17.54	0.263		
ANOVA						
Source of Variation	SS	df	MS	F	P-value	F crit
Rows	10.6736	4	2.6684	0.717234706	0.59240848	3.00691728
Columns	185.5576	4	46.3894	12.46892807	8.5462E-05	3.00691728
Error	59.5264	16	3.7204			
Total	255.7576	24				

### Kết quả và biện luận

$F_R = 0.717 < F_{0.05} = 3.007 \Rightarrow$  Chấp nhận giả thuyết  $H_0$  (Phương pháp)

$F_C = 12.469 > F_{0.05} = 3.007 \Rightarrow$  Bác bỏ giả thuyết  $H_0$  (Dung môi)

Vậy chỉ có dung môi ảnh hưởng đến kết quả chiết xuất.

## C- PHÂN TÍCH PHƯƠNG SAI HAI YẾU TỐ (CÓ LẬP)

### 5.5 Khái niệm thống kê

Tương tự như mô hình phân tích phương sai hai yếu tố (xem tr.67), chỉ là mỗi mức mức ( $A_j, B_j$ ) có sự lặp lại  $k$  lần thí nghiệm mỗi hàng sẽ biểu thị một bản sao của dữ liệu và trong đầu ra sẽ thêm một đại lượng tương tác (interaction term)  $F_1$  giữa hai yếu tố A và B.

**Thí dụ 14:** hàm lượng saponin (mg) của cùng một loại dược liệu được thu hái mùa (khô và mưa: trong mỗi mùa lấy mẫu ba lần - đầu, giữa và cuối) và từ ba miền (nam, trung và bắc) được tóm tắt như sau:

Mùa Thời điểm		Miền		
		Nam	Trung	Bắc
Mùa 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
Mùa 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

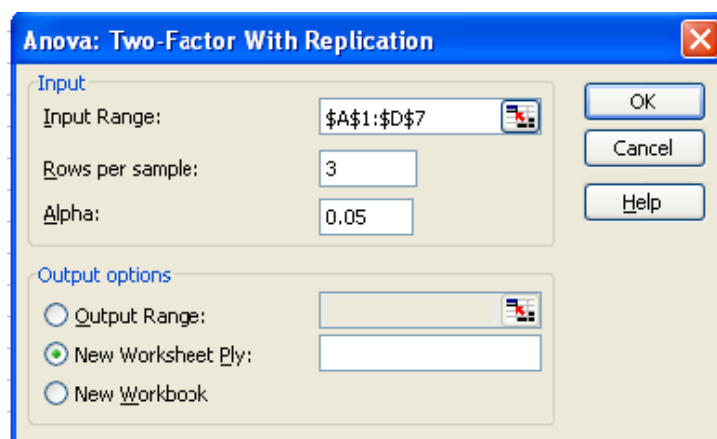
Hãy cho biết hàm lượng saponin có khác nhau theo mùa hay miền? Nếu có thì hai yếu tố mùa và miền có sự tương tác với nhau hay không?

### 5.6 Áp dụng MS-EXCEL

*Nhập dữ liệu vào bảng tính*

	A	B	C	D
1		Nam	Trung	Bắc
2	Mùa khô	2.4	2.1	3.2
3		2.4	2.2	3.2
4		2.5	2.2	3.4
5	Mùa mưa	2.5	2.2	3.4
6		2.5	2.3	3.5
7		2.6	2.3	3.5

*Áp dụng “Anova: Two-Factor With Replication”*



Hình 5.3: Hộp thoại Anova: Two-Factor With Relication



Anova: Two-Factor With Replication						
SUMMARY	Nam	Trung	Bắc	Total		
<i>Mùa khô</i>						
Count	3	3	3	9		
Sum	7.3	6.5	9.8	23.6		
Average	2.433333	2.166667	3.266667	2.622222		
Variance	0.003333	0.003333	0.013333	0.251944		
<i>Mùa mưa</i>						
Count	3	3	3	9		
Sum	7.6	6.8	10.4	24.8		
Average	2.533333	2.266667	3.466667	2.755556		
Variance	0.003333	0.003333	0.003333	0.300278		
<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						
<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	0.001762	4.747225
Columns	4.347778	2	2.173889	434.7778	6.36E-12	3.885294
Interaction	0.01	2	0.005	1	0.396569	3.885294
Within	0.06	12	0.005			
Total	4.497778	17				

### Kết quả và biện luận

$F_R = 16 > F_{0.05} = 4.747 \Rightarrow$  Bác bỏ giả thuyết  $H_0$  (Mùa).

$F_C = 434.778 > F_{0.05} = 3.385 \Rightarrow$  Bác bỏ giả thuyết  $H_0$  (Miền)

$F_1 = 1 < F_{0.05} = 3.385 \Rightarrow$  Chấp nhận giả thuyết  $H_0$  (Mùa x miền).

Vậy hàm lượng Saponin trong dược liệu được khảo sát khác nhau không những theo mùa mà còn theo miền. Tuy nhiên, không có sự tương tác giữa hai yếu tố mùa và miền trên hàm lượng ấy.

## D- PHÂN TÍCH PHƯƠNG SAI BA YẾU TỐ

### 5.7 Khái niệm thống kê

Sự phân tích này được dùng để đánh giá về sự ảnh hưởng của ba yếu tố trên các giá trị quan sát  $\hat{G}$  ( $i = 1, 2, \dots, r$ : yếu tố A;  $j = 1, 2, \dots, r$ : yếu tố B;  $k = 1, 2, \dots, r$ : yếu tố C)

#### Mô hình

Khi nghiên cứu ảnh hưởng của hai yếu tố, mỗi yếu tố có  $n$  mức, thì người ta dùng mô hình vuông la tinh  $n \times n$ , thí dụ như mô hình vuông la tinh  $4 \times 4$ :

B	C	D	A
C	D	A	B
D	A	B	C
A	B	C	D

Mô hình vuông la tinh ba yếu tố được trình bày như sau:

Yếu tố C (T.k. thí dụ:  $T_{...1} = Y_{111} + Y_{421} + Y_{331} + Y_{241}$ )

Yếu tố A	Yếu tố B								T <sub>i...</sub>
	B1		B2		B3		B4		
A1	C1	Y <sub>111</sub>	C2	Y <sub>122</sub>	C3	Y <sub>133</sub>	C4	Y <sub>144</sub>	T <sub>1...</sub>
A2	C2	Y <sub>212</sub>	C3	Y <sub>223</sub>	C4	Y <sub>234</sub>	C1	Y <sub>241</sub>	T <sub>2...</sub>
A3	C3	Y <sub>313</sub>	C4	Y <sub>324</sub>	C1	Y <sub>334</sub>	C2	Y <sub>342</sub>	T <sub>3...</sub>
A4	C4	Y <sub>414</sub>	C1	Y <sub>421</sub>	C2	Y <sub>412</sub>	C3	Y <sub>443</sub>	T <sub>4...</sub>
T <sub>.i.</sub>		T <sub>.1.</sub>		T <sub>.2.</sub>		T <sub>.3.</sub>		T <sub>.4.</sub>	

#### Bảng ANOVA

Nguồn sai số	Bậc tự do	Tổng số bình phương	Bình phương trung bình	Giá trị thống kê
Yếu tố A (Hàng)	$(r-1)$	$SSR = \sum_{i=1}^r \frac{T_{i..}^2}{r} - \frac{T_{...}^2}{r^2}$	$MSR = \frac{SSR}{(r-1)}$	$F_R = \frac{MSR}{MSE}$
Yếu tố B (Cột)	$(r-1)$	$SSC = \sum_{j=1}^r \frac{T_{.j.}^2}{r} - \frac{T_{...}^2}{r^2}$	$MSC = \frac{SSC}{(r-1)}$	$F_C = \frac{MSC}{MSE}$
Yếu tố C	$(r-1)$	$SSF = \sum_{k=1}^r \frac{T_{...k}^2}{r} - \frac{T_{...}^2}{r^2}$	$MSF = \frac{SSF}{(r-1)}$	$F = \frac{MSF}{MSE}$
Sai số	$(r-1)(r-2)$	$SSE = SST - (SSF + SSR + SSC)$	$MSE = \frac{SSE}{(r-1)(r-2)}$	
Tổng cộng	$(r^2-1)$	$SST = \sum_i \sum_j \sum_k Y_{ijk}^2 - \frac{T_{...}^2}{r^2}$		

#### Trắc nghiệm

\* Giả thiết:

$H_0 : \mu_1 = \mu_2 = \dots \mu_k \Leftrightarrow$  « Các giá trị trung bình bằng nhau »

$H_1 : \mu_i \neq \mu_j \Leftrightarrow$  « Có ít nhất hai giá trị trung bình khác nhau »

\* Giá trị thống kê:  $\hat{G}$  và  $\hat{G}$

\* Biện luận:

- Nếu  $\hat{G}$  ( Chấp nhận  $H_0$  (Yếu tố A)

- Nếu  $\hat{G}$  ( Chấp nhận  $H_0$  (Yếu tố B)

- Nếu  $\hat{G}$  ( Chấp nhận  $H_0$  (Yếu tố C).

## 5.8 Áp dụng MS-EXCEL

**Thí dụ 15:** Hiệu suất phần trăm (%) của một phản ứng hóa học được nghiên cứu theo ba yếu tố: pH (A), nhiệt độ (B) và chất xúc tác (C) được trình bày trong bảng sau:

Yếu tố A	Yếu tố B							
	B1		B2		B3		B4	
A1	C1	9	C2	14	C3	16	C4	12
A2	C2	12	C3	15	C4	12	C1	10
A3	C3	13	C4	14	C1	11	C2	14
A4	C4	10	C1	11	C2	13	C3	13

Hãy đánh giá về ảnh hưởng của các yếu tố trên hiệu suất phản ứng?

**Nhập dữ liệu vào bảng tính**

1	A	B	C	D	E	F	G	H	I	J	K	L	M
2		B1	B2	B3	B4								
3	A1	9	14	16	12								
4	A2	12	15	12	10								
5	A3	13	14	11	14								
6	A4	10	11	13	13								
7													
8	Ti..	51	49	52	47	SUMSQTi..	9915	SSR	3.69	MSR	1.2292	FR	3.106
9	T.j.	44	54	52	49	SUMSQT.j.	9957	SSC	14.2	MSC	4.7292	FC	11.95
10	T..k	41	53	57	48	SUMSQT..k	10043	SSF	35.7	MSF	11.896	F	30.06
11	T...	199				SQT...	39601	SSE	2.38	MSE	0.3958		
12						SUMSQ	2531	STT	55.9				
						Yijk							

Thiết lập các biểu thức và tính các giá trị thống kê

\* Tính các giá trị Ti.... T.j.. T..k và T...

- Các giá trị Ti..

Chọn ô B7 và nhập biểu thức =SUM(B2:E2)

Chọn ô C7 và nhập biểu thức =SUM(B3:E3)

Chọn ô D7 và nhập biểu thức =SUM(B4:E4)

Chọn ô E7 và nhập biểu thức =SUM(B5:E5)

Các giá trị T.j.

Chọn ô B8 và nhập biểu thức =SUM(B2:B5)

Dùng con trỏ kéo k1i hiệu tự điền từ ô B8 đến ô E8

**- Các giá trị T..k**

Chọn ô B9 và nhập biểu thức =SUM(B2.C5.D4.E3)

Chọn ô C9 và nhập biểu thức =SUM(B3.C2.D5.E4)

Chọn ô D9 và nhập biểu thức =SUM(B4.C3.D2.E5)

Chọn ô E9 và nhập biểu thức =SUM(B5.C4.D3.E2)

**Giá trị T...**

Chọn ô B10 và nhập biểu thức =SUM(B2:E5)

**\* Tính các giá trị  $\hat{G}$  và  $\hat{G}$**

**- Các giá trị  $\hat{G}$  và  $\hat{G}$**

Chọn ô G7 và nhập biểu thức =SUMSQ(B7:B7)

Dùng con trỏ kéo kí hiệu tự điền từ ô G7 đến ô G9

**- Giá trị  $\hat{G}$**

Chọn ô G10 và nhập biểu thức =POWER(B10.2)

**- Giá trị  $\hat{G}$**

Chọn ô G11 và nhập biểu thức =SUMSQ(B2:E5)

**\* Tính các giá trị SSR. SSC. SSF. SST và SSE**

**- Các giá trị SSR. SSC và SSF**

Chọn ô I7 và nhập biểu thức =G7/4-39601/POWER(4.2)

Dùng con trỏ kéo kí hiệu tự điền từ ô I7 đến ô I9

**- Giá trị SST**

Chọn ô I11 và nhập biểu thức =G11-G10/POWER(4.2)

**- Giá trị SSE**

Chọn ô I10 và nhập biểu thức =I11-SUM(I7:I9)

**\* Tính các giá trị MSR. MSC. MSF và MSE**

**- Các giá trị MSR. MSC và MSF**

**- Giá trị SST**

Chọn ô K7 và nhập biểu thức =I7/(4-1)

Dùng con trỏ kéo kí hiệu tự điền từ ô K7 đến ô K9.

**- Giá trị MSE**

Chọn ô K10 và nhập biểu thức =I10/((4-1)\*(4-2)).

**\* Tính các giá trị  $\bar{G}$  và  $F$** 

Chọn ô M7 và nhập biểu thức = K7/0.3958

Dùng con trỏ kéo kí hiệu tự điền từ ô M7 đến ô M9.

**Kết quả và biện luận**

$F_R = 3,10 < F_{0,05}(3,6) = 4,76 \Rightarrow$  chấp nhận  $H_0$  (pH).

$F_C = 11,95 > F_{0,05}(3,6) = 4,76 \Rightarrow$  Bác bỏ  $H_0$  (nhiệt độ)

$F = 30,05 > F_{0,05}(3,6) = 4,76 \Rightarrow$  bác bỏ  $H_0$  (chất xúc tác)

Vậy chỉ có nhiệt độ và chất xúc tác gây ảnh hưởng đến hiệu suất.