

CHƯƠNG VI

KIỂM ĐỊNH GIẢ THUYẾT THỐNG KÊ

Đây là bài toán thống kê rất hay gặp trong thực nghiệm. Ở đây đề cập đến việc kiểm định (so sánh) hai số trung bình của một đặc tính định lượng hay hai xác suất (hai tỷ lệ) của một đặc tính định tính của quần thể.

6.1. Những khái niệm chung và ý nghĩa

Trong nghiên cứu thường phải so sánh các tham số thống kê như số trung bình, phương sai, xác suất của một mẫu với một tiêu chuẩn cho trước nào đó, hoặc 2 mẫu với nhau hay nhiều mẫu với nhau. Thông thường các tham số có sự khác nhau (khác nhau về số học), nhưng ta lại cần xem xét sự sai khác này có rõ ràng hay không? ở mức độ nào?

Nếu chúng chỉ khác nhau trong phạm vi ngẫu nhiên thì sự khác nhau này được coi như không đáng kể (không có ý nghĩa). Nếu chúng khác nhau ngoài phạm vi ngẫu nhiên thì kết luận sự khác nhau ấy là do tác động của nhân tố thí nghiệm (ngoài tác động của nhân tố ngẫu nhiên). Đây là vấn đề mà các nhà sinh học nông nghiệp cần quan tâm.

Để kiểm định người ta dùng các kết quả thực nghiệm quan sát ở mẫu với việc vận dụng công cụ toán học là lý thuyết xác suất để kiểm tra những giả thiết đã cho. Nếu tài liệu thực nghiệm phù hợp với giả thiết thì giả thiết được chấp nhận. Ngược lại thì giả thiết bị bác bỏ. Sự phù hợp mà ta nói ở đây không phải là tuyệt đối, mà chỉ là phù hợp theo một tiêu chuẩn nào đó xác định trước đủ thoả mãn những yêu cầu của thực tiễn.

Thí dụ: Trong nông học người ta thường so sánh (hay kiểm định) sự sinh trưởng, phát triển, diễn biến sâu bệnh hại cây trồng cũng như các chỉ tiêu năng suất được gieo trồng bằng những biện pháp kỹ thuật khác nhau để xem chúng có ảnh hưởng thực sự đến các chỉ tiêu nghiên cứu hay không?

Trong chương này chúng tôi sẽ tập trung đề cập đến một số tiêu chuẩn thống kê được dùng để kiểm định trong các trường hợp cụ thể.

6.2. Trường hợp hai mẫu độc lập

Mẫu độc lập hay thí nghiệm độc lập là những khái niệm tương đối. Theo nghĩa rộng người ta gọi mẫu độc lập hay thí nghiệm độc lập nếu một quá trình thực nghiệm nào đó được thiết kế một cách độc lập với những thí nghiệm khác.

6.2.1. Tiêu chuẩn U của phân phối tiêu chuẩn

Nếu trong trường hợp kiểu phân phối lý thuyết đặc trưng cho 2 kết quả (2 mẫu) nghiên cứu chưa biết thì yêu cầu dung lượng mẫu lấy phải được coi là đủ lớn ($n_1 > 30$ và $n_2 > 30$). Theo luật số lớn thì trong trường hợp mẫu lớn, phân phối xác suất của số trung bình mẫu $\bar{X}$

xấp xỉ luật chuẩn với kỳ vọng $M\bar{X} = \mu$ và phương sai $D\bar{X} = \frac{s^2}{n}$

$$\text{Nhu vậy } \bar{X}_1 \sim N\left(\mu_1, \frac{\sigma_1^2}{n_1}\right)$$

$$\bar{X}_2 \sim N\left(\mu_2, \frac{\sigma_2^2}{n_2}\right)$$

Giả thiết H_0 : $\mu_1 = \mu_2$ hay $\mu_1 - \mu_2 = 0$

Đối thiết H_1 : $\mu_1 \neq \mu_2$ hay $\mu_1 - \mu_2 \neq 0$ (6.1)

$$\overline{X}_1 - \overline{X}_2 \sim N\left(\mu_1 - \mu_2, \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}\right) \quad (6.2)$$

Được kiểm định bằng tiêu chuẩn u của phân phối tiêu chuẩn với mức ý nghĩa α , tính giá trị thực nghiệm như sau:

$$u_m = \frac{|\overline{x}_1 - \overline{x}_2|}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \quad (6.3)$$

Nếu phương sai của 2 tổng thể không được biết trước và dung lượng mẫu đủ lớn thì có thể thay một cách gần đúng phương sai tổng thể bằng phương sai mẫu. Có nghĩa là

$$\sigma_1^2 \approx s_1^2 \quad \text{và} \quad \sigma_2^2 \approx s_2^2$$

Lúc này tiêu chuẩn phù hợp như sau:

$$u_{t.n} = \frac{|\overline{x}_1 - \overline{x}_2|}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} \quad (6.4)$$

Nếu như $|u_{t.n}| < u_\alpha$ tra ở bảng Φ (bảng 1 phụ lục) với mức ý nghĩa α thì giả thiết H_0 được chấp nhận nghĩa là hai trung bình của hai mẫu bằng nhau. Ngược lại nếu $|u_{t.n}| \geq u_\alpha$ thì giả thiết bị bác bỏ nghĩa là hai trung bình của hai mẫu là khác nhau.

Thí dụ

Đo chiều cao cây cuối cùng của 2 giống lúa mới có kết quả như sau:

Giống I: Đo $n = 42$ khóm có chiều cao trung bình $\overline{x}_1 = 95,2cm$

Độ lệch chuẩn về chiều cao là $s_1 = 3,2 cm$.

Giống II: Đo $n = 40$ khóm có chiều cao trung bình $\overline{x}_1 = 98,5cm$

Độ lệch chuẩn tương ứng $s_2 = 3,4 cm$.

Hỏi chiều cao cây của hai giống có khác nhau hay không với mức ý nghĩa $\alpha = 0,05$

Vì hai dung lượng mẫu n_1 và n_2 lấy từ hai giống lúa nghiên cứu là đủ lớn, và không biết trước được hai phương sai tổng thể. Nên ta có thể dùng tiêu chuẩn u của phân phối chuẩn để kiểm định.

Giả thiết H_0 : $\mu_1 = \mu_2$ hay $\mu_1 - \mu_2 = 0$

Đối thiết H_1 : $\mu_1 \neq \mu_2$ hay $\mu_1 - \mu_2 \neq 0$

Để kiểm định giả thiết H_0 ta áp dụng biểu thức (6.4).

$$u_{t,n} = \frac{|\bar{x}_1 - \bar{x}_2|}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} = \frac{|95,2 - 98,5|}{\sqrt{\frac{3,2^2}{42} + \frac{3,4^2}{40}}} = |4,52|$$

Với $\alpha = 0,05$ $u_\alpha = 1,96$

Ở đây $|u_m| > u_{0,05} (|4,52| > 1,96)$

Nên ta bác bỏ giả thiết H_0 và chấp nhận đối thiết H_1 là giống khác nhau thì chiều cao khác nhau rõ rệt.

6.2.2. Tiêu chuẩn t của phân phối Student

Tiêu chuẩn này được áp dụng trong trường hợp luật phân phối của hai tổng thể mà đại diện là 2 mẫu có phân phối chuẩn và phương sai của hai tổng thể được coi là bằng nhau. Nếu thoả mãn hai điều kiện này thì có thể kiểm tra.

Giả thiết H_0 : $\mu_1 = \mu_2$ hay $\mu_1 - \mu_2 = 0$

Đối thiết H_1 : $\mu_1 \neq \mu_2$ hay $\mu_1 - \mu_2 \neq 0$ bằng tiêu chuẩn T của Student như sau:

$$T_{t,n} = \frac{|\bar{x}_1 - \bar{x}_2|}{\sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2} \left(\frac{1}{n_1} + \frac{1}{n_2} \right)}} \quad (6.5)$$

Trong đó: $\bar{x}_1$ và $\bar{x}_2$ là trung bình của 2 mẫu là mẫu 1 và mẫu 2.

s_1^2 và s_2^2 là phương sai của mẫu 1 và mẫu 2.

n_1 và n_2 là dung lượng quan sát của 2 mẫu.

Thường trường hợp này được áp dụng khi n_1 và n_2 là không đủ lớn

$n_1 < 30$ và $n_2 < 30$ hoặc $n_1 > 30$ và $n_2 < 30$ hoặc $n_2 > 30$ và $n_1 < 30$

Nếu $|T_{t,n}| < t_{(\alpha, n_1 + n_2 - 2)}$ tra bảng t với $n_1 + n_2 - 2$ bậc tự do (phụ lục bảng 4) thì giả thiết H_0 được chấp nhận, nghĩa là trung bình của hai mẫu bằng nhau. Ngược lại nếu như $|T_{t,n}| > t_{(\alpha, n_1 + n_2 - 2)}$ tra bảng với bậc tự do $= n_1 + n_2 - 2$ thì giả thiết H_0 bị bác bỏ. Nghĩa là chấp nhận đối thuyết H_1 trung bình hai mẫu là khác nhau.

Thí dụ

So sánh năng suất của hai giống cà chua vụ xuân hè

Giống số 6: theo dõi 16 địa điểm năng suất trung bình hạt 30,6 tấn/ha.

Độ lệch chuẩn về năng suất 4,5 tấn/ha.

Giống số 204 A theo dõi ở 19 địa điểm, năng suất bình quân đạt 27,0 tấn/ha. Độ lệch chuẩn năng suất là 4,0 tấn/ha.

Biết rằng phân phối về năng suất của cà chua là phân phối chuẩn và hai phương sai lý thuyết được coi là bằng nhau.

Hãy cho biết là trung bình của hai giống trên có khác nhau hay không ở mức ý nghĩa $\alpha = 0,05$ và 0,01.

Vì hai tổng thể đã đáp ứng các điều kiện nêu ra nên có thể áp dụng tiêu chuẩn t của phân phối Student để kiểm định giả thiết

Giả thiết H_0 : $\mu_1 = \mu_2$ hay $\mu_1 - \mu_2 = 0$

Đối thiết H_1 : $\mu_1 \neq \mu_2$ hay $\mu_1 - \mu_2 \neq 0$

Theo (6.5)

$$T_{t,n} = \frac{|\bar{x}_1 - \bar{x}_2|}{\sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2} \left(\frac{1}{n_1} + \frac{1}{n_2} \right)}}$$

Thay vào được

$$T_{t,n} = \frac{|30,6 - 27,0|}{\sqrt{\frac{(16 - 1) \cdot 4,5^2 + (19 - 1) \cdot 4,0^2}{16 + 19 - 2} \left(\frac{1}{16} + \frac{1}{19} \right)}} = 2,064$$

$t_{(0,05;33)} = 2,04$, như vậy $(|T_{t,n}| > t_{0,05;33})$ nên giả thiết H_0 bị bác bỏ và chấp nhận H_1 với câu trả lời "hai giống cà chua nói trên có năng suất trung bình khác nhau".

Với $\alpha = 0,01$ thì $t_{(0,01;33)} = 2,75$

Vì $2,064 < 2,75$ nên chấp nhận giả thiết H_0 có nghĩa là "hai giống cà chua nói trên có năng suất trung bình bằng nhau".

Trong một số trường hợp khi áp dụng tiêu chuẩn t của Student để so sánh 2 mẫu độc lập, điều kiện về luật phân phối chuẩn của tổng thể được thoả mãn, nhưng 2 phương sai của tổng thể không được biết trước.

Khi đó phải kiểm định bằng tiêu chuẩn F của Fisher về sự bằng nhau của hai phương sai (sẽ nêu ở mục 6.5).

Nếu hai phương sai bằng nhau thì áp dụng kiểm định như biểu thức (6.5).

Nếu hai phương sai được coi là khác nhau thì biểu thức áp dụng cho tiêu chuẩn t của Student như sau:

$$T_{t,n} = \frac{|\bar{x}_1 - \bar{x}_2|}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} \quad (6.6)$$

Nhưng $|T_{t,n}|$ sẽ được so sánh với giá trị t lý thuyết ở mức ý nghĩa α tính như sau:

$$t_{\alpha}^* = \frac{\left[t_{(\alpha, df_1=n_1-1)} \frac{s_1^2}{n_1} \right] + \left[t_{(\alpha, df_2=n_2-1)} \frac{s_2^2}{n_2} \right]}{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} \quad (6.7)$$

Nếu như $|T_{t,n} < t^*|$ thì chấp nhận H_0 (hai trung bình là bằng nhau)

Nếu như $|T_{t,n} > t^*|$ bác bỏ giả thiết H_0 thì chấp nhận H_1 (hai trung bình là khác nhau).

Thí dụ

Phân tích hàm lượng đường tổng số (%) của hai giống cà chua vụ xuân hè.

- Giống MV1 phân tích ở 6 mẫu có hàm lượng đường tổng số đạt 3,09% và phương sai về đường tổng số là 0,75 %.

- Giống mới phân tích ở $n_2 = 8$ mẫu có giá trị trung bình của đường tổng số đạt 2,79%. Phương sai có giá trị 0,10 %.

Hãy cho hàm lượng đường tổng số của hai giống nêu trên có khác nhau hay không ?

Cho hai phương sai là không bằng nhau với mức ý nghĩa $\alpha = 0,05$.

Do hai phương sai không bằng nhau nên áp dụng biểu thức (6.6)

$$T_{t,n} = \frac{|\bar{x}_1 - \bar{x}_2|}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} \text{ thay số vào ta có } T_{t,n} = \frac{|3,09 - 2,79|}{\sqrt{\frac{0,75}{6} + \frac{0,16}{8}}} = |0,798|$$

Áp dụng biểu thức (6.7) tính $t_{0,05}^*$ như sau:

$$t_{\alpha}^* = \frac{\left[t_{(\alpha, df_1=n_1-1)} \frac{s_1^2}{n_1} \right] + \left[t_{(\alpha, df_2=n_2-1)} \frac{s_2^2}{n_2} \right]}{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

Thay số vào

$$t_{0,05}^* = \frac{\left(2,75 \times \frac{0,75}{6} \right) + \left(2,36 \times \frac{0,10}{8} \right)}{\frac{0,75}{6} + \frac{0,10}{8}} = 2,54$$

Như vậy, $|T_{t,n} < t^*|$ ($0,798 < 2,54$) nên chấp nhận giả thiết H_0 "hai giống cà chua nói trên có hàm lượng đường tổng số như nhau"

6.3. Trường hợp hai mẫu theo cặp

6.3.1. Khái niệm về mẫu theo cặp

Thực tiễn của nghiên cứu nông học nói riêng và nghiên cứu trong nông nghiệp cũng như sinh học nói chung, ngoài những mẫu độc lập như đã nêu ở trên, nhà khoa học khi nghiên cứu còn gặp một loạt các mẫu quan sát không độc lập mà lại có xếp theo cặp (đôi)

Thí dụ: khi đo chiều cao cây ta dùng hai loại thước đo để đo cho cùng một mẫu. Hoặc xác định khối lượng 1000 hạt của mẫu bằng hai chiếc cân khác nhau. Hay hai kỹ thuật viên A và B đếm khuẩn lạc của n hộp petri. Như vậy, các mẫu trên đều được thực hiện 2 lần quan sát, kết quả được xác định theo cặp.

Những kết quả như vậy sẽ được kiểm định theo phương pháp cặp. Do vậy, người làm công tác thống kê sẽ vận dụng các tiêu chuẩn khác cho phù hợp. Dưới đây chúng tôi trình bày tiêu chuẩn thông thường nhất qua thí dụ sau:

6.3.2. Tiêu chuẩn t của Student

Giả thiết H_0 : $\mu_x = \mu_y$ hay $\mu_x - \mu_y = 0$

Đối thiết H_1 : $\mu_x \neq \mu_y$ hay $\mu_x - \mu_y \neq 0$

Tính các hiệu sai $d_i = y_i - x_i$ (hiệu của hai mẫu) sau đó tính độ lệch chuẩn của hiệu sai theo biểu thức sau:

$$|T_{t,n}| = \frac{\bar{d}}{s_{\bar{d}}} = \frac{|\bar{d}|}{\sqrt{\frac{\sum d_i^2 - \frac{(\sum d_i)^2}{n}}{n(n-1)}}} \quad (6.8)$$

$$\text{với } \bar{d} = \frac{\sum d_i}{n} \quad (6.9)$$

$$s_d^2 = \frac{\sum d_i^2 - \frac{(\sum d_i)^2}{n}}{(n-1)} \quad (6.10)$$

Sai số chuẩn $s_{\bar{d}} = \frac{s_d}{\sqrt{n}}$ (Độ lệch chuẩn hiệu sai bình quân)

$|T_{t,n}| < t_{(\alpha,n-1)}$ chấp nhận H_0

$|T_{t,n}| > t_{(\alpha,n-1)}$ chấp nhận H_1

Thí dụ

Đo đường vanh của 18 cây cao su bằng hai loại thước khác nhau kết quả được ghi lại trong bảng 6.1 sau. Có sự sai khác rõ rệt giữa 2 loại thước đo hay không ?

Bảng 6.1. So sánh hai mẫu theo cặp bằng tiêu chuẩn t

TT	x_i của thước 1	y_i của thước 2	$d_i = x_i - y_i$	d_i^2
1	29,7	29,0	0,7	0,49
2	34,9	34,5	0,4	0,16
3	37,0	37,5	- 0,5	0,25
4	28,5	28,0	0,5	0,25
5	26,9	26,5	0,4	0,16
6	27,0	27,1	- 0,1	0,01
7	28,7	28,5	0,2	0,04
8	31,2	31,0	0,2	0,04
9	31,5	31,0	0,5	0,25
10	33,7	33,5	0,2	0,04

11	34,0	33,7	0,3	0,09
12	35,9	36,0	-0,1	0,01
13	41,2	41,5	-0,3	0,09
14	29,5	29,0	0,5	0,25
15	30,7	30,5	0,2	0,04
16	33,5	33,7	-0,2	0,04
17	36,2	36,0	0,2	0,04
18	37,5	37,5	0	0
Tổng			$\Sigma d_i = 3,1$	$\Sigma d_i^2 = 2,25$

$$s_{\bar{d}} = \sqrt{\frac{\sum di^2 - (\sum di)^2}{n(n-1)}} = \sqrt{\frac{2,25 - \frac{(3,1)^2}{18}}{18(18-1)}} = 0,0749$$

Thay số vào biểu thức (6.8)

$$|T_{t,n}| = \frac{0,01722}{0,0749} = 2,30$$

Tra bảng t của Student với $\alpha = 0,05$ bậc tự do $df = 17$ ta được giá trị $t_{(0,05;17)} = 2,11$

Vì $|T_{t,n}| > t_{(0,05;17)}$ nên hai thước đo khác nhau thì kết quả trung bình khác nhau rõ rệt.

6.4. Kiểm định tính độc lập (kiểm định tính thuần nhất của các mẫu về chất)

Đây là trường hợp áp dụng cho số liệu nghiên cứu mang tính định tính, hay còn gọi là kiểm định phi tham số.

6.4.1. So sánh 2 tỷ lệ (hay gọi là kiểm định 2 xác suất)

Giả sử có 2 tổng thể (2 công thức), công thức I có N_1 phân tử, trong đó có M_1 phân tử có đặc tính A và $N_1 - M_1$ phân tử không có đặc tính A.

Công thức II có N_2 phân tử, trong đó có M_2 phân tử có đặc tính A và tương tự $N_2 - M_2$ phân tử không có đặc tính A.

Từ 2 công thức rút mẫu ngẫu nhiên với dung lượng n_1 và n_2 . Như vậy, tần suất của A ở các mẫu lần lượt là

$$f_1 = \frac{m_1}{n_1} \quad \text{và} \quad f_2 = \frac{m_2}{n_2}$$

Dựa vào 2 kết quả này để kiểm định giả thiết H_0 với nội dung là

$$H_0: p_1 = p_2 \quad (p_1 - p_2 = 0) \tag{6.11}$$

Ở đây p_1 là xác suất của các cá thể A trong công thức I

p_2 là xác suất của các cá thể A trong công thức II

Để có thể chấp nhận hay bác bỏ giả thiết H_0 ta phải xem xét cụ thể qua các trường hợp sau:

6.4.1.1. Trường hợp 2 dung lượng mẫu đều lớn

$n_1 > 100$ và $n_2 > 100$ và P_0 xa 0 và 1

Trường hợp này áp dụng tiêu chuẩn U của hàm phân phối chuẩn như sau:

$$|u_{t,n}| = \frac{|f_1 - f_2|}{\sqrt{\frac{f_1(1-f_1)}{n_1} + \frac{f_2(1-f_2)}{n_2}}} \quad (6.12)$$

$|u_{t,n}| < u_\alpha$ chấp nhận giả thiết có nghĩa là coi hai xác suất p_1 và p_2 như nhau.

$|u_{t,n}| \geq u_\alpha$ bác bỏ giả thiết H_0 nghĩa là coi $p_1 \neq p_2$

Thí dụ

Nghiên cứu hiện tượng sâu đục quả với 2 giống cà chua được trồng vụ xuân hè tại đồng bằng sông Hồng. Giống cà chua số 6 trong $n_1 = 300$ quả có 36 quả bị sâu đục.

Giống số 48 trong $n_2 = 240$ quả có 38 quả bị sâu đục.

Hỏi: tỷ lệ sâu đục quả của 2 giống cà chua có khác nhau hay không? (cho $\alpha = 0,05$)

Tỷ lệ bị sâu ở giống số 6 là $f_1 = \frac{36}{300} = 0,120$ hay 12,0%

Tỷ lệ bị sâu ở giống số 48 là $f_2 = \frac{38}{240} = 0,158$ hay 15,8%

Thay số vào biểu thức (6.12)

$$|u_{t,n}| = \frac{f_1 - f_2}{\sqrt{\frac{f_1(1-f_1)}{n_1} + \frac{f_2(1-f_2)}{n_2}}} = \frac{0,120 - 0,158}{\sqrt{\frac{0,12 \times 0,88}{300} + \frac{0,158}{240}}} = 1,27$$

$$u_{0,05} = 1,96$$

Như vậy $|u_{t,n}| < 1,96$ chấp nhận giả thiết H_0 có nghĩa là ở mức ý nghĩa 0,05 "tỷ lệ sâu đục quả của 2 giống cà chua như nhau".

6.4.1.2. Trường hợp n_1 hoặc $n_2 < 100$ hoặc cả n_1 và n_2 nhỏ hơn 100)

Trường hợp này ta phải sử dụng hàm phân phối χ^2 (khi bình phương) để kiểm định. Có thể tính được $\chi_{t,n}^2$ bằng một biểu thức sau:

$$\chi_{t,n}^2 = \sum_{i=1}^{2 \times 2} \frac{|(m_{ij} - M_{ij}) - 0,5|^2}{M_{ij}} \quad (6.13)$$

Ở đây m_{ij} là tần số thực nghiệm của đặc tính j trong công thức thứ i

M_{ij} là tần số lý thuyết của đặc tính j ở công thức thứ i

0,5 là hệ số hiệu chỉnh YATES.

Nếu $\chi^2_{t,n} < \chi^2_{(\alpha,1)}$ giá trị χ^2 ở mức nghĩa α và bậc tự do $df = 1$ thì chấp nhận giả thiết H_0 $p_1 = p_2$ (2 tỷ lệ đem kiểm định là như nhau).

Ngược lại nếu $\chi^2_{t,n} \geq \chi^2_{(\alpha,1)}$ bác bỏ giả thiết H_0 và coi hai tỷ lệ đem kiểm định là khác nhau.

Thí dụ

Khảo sát hai phương pháp trừ bệnh sương mai cho cà chua có kết quả như sau:

Phương pháp I: Phun cho $n_1 = 70$ cây có 52 cây không có bệnh.

Phương pháp II: Phun cho $n_2 = 50$ cây có 30 cây không có bệnh.

Vậy phương pháp trừ sương mai cho cà chua khác nhau thì hiệu lực có khác nhau hay không?

Các tần số lý thuyết tính như sau

Nếu giả thiết H_0 : $p_1 = p_2$ là đúng trong tổng số $70 + 50 = 120$ cây có $52 + 30 = 82$ cây không bị bệnh, vậy có thể coi số cây không bị bệnh của phương pháp I là

$$\frac{82}{120} \times 70 = 47,83 \approx 48 \text{ cây}$$

Số cây vẫn bị sương mai của phương pháp I là $70 - 48 \text{ cây} = 22 \text{ cây}$.

Phương pháp II có số cây không bệnh là

$$\frac{82}{120} \times 50 = 34,16 \approx 34 \text{ cây}$$

Số cây vẫn bị sương mai của phương pháp II là $50 - 34 \text{ cây} = 16 \text{ cây}$.

Như vậy, thay vào công thức (6.13) để tính $\chi^2_{t,n}$ như sau:

$$\chi^2_{t,n} = \frac{|(52 - 48) - 0,5|^2}{48} + \frac{|(18 - 22) - 0,5|^2}{22} + \frac{|(30 - 34) - 0,5|^2}{34} + \frac{|(20 - 16) - 0,5|^2}{16} = 1,94$$

Tra bảng χ^2 với $\alpha = 0,05$ và $df = 1$ được $\chi^2_{(0,05;1)} = 3,84$ (Bảng 3 phụ lục)

Như vậy, $\chi^2_{y,n} < \chi^2_{(0,05;1)}$ nên chấp nhận giả thiết H_0 là hai phương pháp trừ bệnh sương mai khác nhau thì kết quả như nhau ở mức $\alpha = 0,05$.

6.4.1.3. Trường hợp hai sự kiện hiếm (hay xác suất bé)

Trường hợp này biến đổi để đưa về biến có phân phối chuẩn

Công thức biến đổi như sau:

$$\varphi = 2 \frac{\pi}{180} \arcsin \sqrt{f}$$

Ở đây $\pi = 3,1416$; f là tần suất của mẫu

Có bảng tính sẵn đối với $f = 0,00001$ đến $0,0099$

(Bảng 9 phụ lục)

Sau khi đã biến đổi ta áp dụng công thức sau

$$|u_m| = \frac{\varphi_1 - \varphi_2}{\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \quad (6.14)$$

$|u_{t,n}|$ được so sánh với u_α nếu $|u_{t,n}| < u_\alpha$ chấp nhận giả thiết H_0 (hai tỷ lệ so sánh là như nhau)

Ngược lại nếu $|u_{t,n}| \geq u_\alpha$ bác bỏ H_0 (hai tỷ lệ đó là khác nhau).

Thí dụ: Có 2 liều lượng xử lý gây đột biến kiểu hình ở đậu tương bằng phương pháp vật lý. Liều lượng I xử lý 5000 hạt thấy có 6 cây dị hình. Liều lượng II xử lý 10.000 hạt thấy có 9 cây dị hình.

Hãy cho biết liều lượng xử lý khác nhau có dẫn tới phản ứng gây đột biến kiểu hình khác nhau hay không với mức ý nghĩa $\alpha = 0,05$.

$n_1 = 5000$ và như vậy xác suất của đột biến là

$$f_1 = \frac{6}{5000} = 0,0012$$

$n_2 = 10.000$ xác suất gây đột biến là

$$f_2 = \frac{9}{10.000} = 0,0009$$

Tương tự có các giá trị tra bảng (9) phần phụ lục là:

$$\varphi_1 = 0,0693 \text{ và } \varphi_2 = 0,0600$$

Áp dụng biểu thức (6.14)

$$|u_m| = \frac{\kappa_2}{\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} = \frac{0,0693 - 0,0600}{\sqrt{\frac{1}{5000} + \frac{1}{10000}}} = 0,537$$

$$u_{0,05} = 1,96$$

$$|u_{t,n}| < u_{0,05}$$

Chấp nhận giả thiết H_0 "hai liều lượng xử lý khác nhau thì hiệu quả gây đột biến kiểu hình như nhau"

6.4.2. Kiểm tra tính thuần nhất của nhiều mẫu về chất (kiểm định nhiều xác suất)

Trường hợp có k công thức và các kết quả quan sát ở mẫu được chia thành r cấp (chất lượng) khác nhau thì giả thiết H_0 được đưa ra là:

H_0 : các mẫu về chất là thuần nhất và tiêu chuẩn kiểm định phù hợp là χ^2 , trong đó giá trị $\chi_{t,n}^2$ (khi bình phương thực nghiệm) tính như sau:

$$\chi_m^2 = N \left[\sum_{i,j} \frac{m_{ij}^2}{t_j \times n_i} - 1 \right] \quad (6.15)$$

Trong đó:

k: là số công thức

r : là số cấp chất lượng

m_{ij} : là tần số quan sát thực nghiệm ở công thức thứ i và đặc trưng định tính (tính chất) thứ j .

t_j : tổng tần số quan sát của các công thức đối với cấp j

n_i : dung lượng mẫu của công thức thứ i

N là tổng số các quan sát; $N = n_1 + n_2 + \dots + n_k = t_1 + t_2 + \dots + t_r$

Nếu $\chi^2_{t,n} < \chi^2_{(\alpha, df)}$ với bậc tự do $df = (k - 1)(r - 1)$ thì chấp nhận giả thiết H_0 có nghĩa là các công thức khác nhau (các mẫu) đồng nhất về chất. Còn nếu $\chi^2_{t,n} \geq \chi^2_{(\alpha, df)}$ thì bác bỏ giả thiết H_0 . Như vậy phải chấp nhận các công thức khác nhau thì chất lượng khác nhau (không thuần nhất).

Số liệu tổng quát được trình bày trong bảng sau (bảng 6.2)

Bảng 6.2. Sắp xếp các số liệu thực nghiệm (tần số m_{ij})

Công thức Cấp	I	II	...	i	...	k	T_i
1	$m_{1,1}$	$m_{2,1}$	...	$m_{i,1}$	...	$m_{k,1}$	T_1
2	$m_{1,2}$	$m_{2,2}$	...	$m_{i,2}$	...	$m_{k,2}$	T_2
.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.
j	$m_{1,j}$	$m_{2,j}$	...	$m_{i,j}$	...	$m_{k,j}$	T_j
.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.
r	$m_{1,r}$	$m_{2,r}$	...	$m_{1,r}$	...	$m_{k,r}$	T_r
n_i	n_1	n_2	...	n_i	...	n_k	N

Thí dụ

Có 3 mật độ trồng cà phê với I, II, III với cùng biện pháp kỹ thuật, cùng tuổi cây. Theo dõi mức độ bệnh gỉ sắt hại cà phê theo 3 cấp (nhẹ, trung bình và nặng). Kết quả quan sát như sau:

Bảng 6.3. Các tần số thực nghiệm

Mật độ Chất lượng	I	II	III	Tổng số
Nhẹ	80	65	60	205
Trung bình	60	75	130	265
Nặng	70	80	70	220
Tổng số	210	220	260	690

Hỏi mật độ khác nhau thì mức độ nhiễm bệnh gi sắt có khác nhau hay không ở mức ý nghĩa $\alpha = 0,05$.

Giả thiết H_0 : mật độ trồng cà phê khác nhau nhưng mức độ nhiễm gi sắt là như nhau

Để chấp nhận giả thiết H_0 hoặc bác bỏ H_0 ta áp dụng biểu thức (6.15) qua tiêu chuẩn χ^2

$$\chi_m^2 = N \left[\sum_{i,j=1}^{3 \times 3} \frac{m_{ij}^2}{t_j \times n_i} - 1 \right]$$

thay số vào ta có:

$$\chi_{t.n}^2 = 690 \left(\frac{80^2}{205 \times 210} + \frac{65^2}{205 \times 220} + \frac{60^2}{205 \times 260} + \frac{60^2}{265 \times 210} + \frac{75^2}{265 \times 220} + \frac{130^2}{265 \times 210} + \frac{70^2}{220 \times 210} + \frac{80^2}{220 \times 220} + \frac{70^2}{220 \times 260} \right) - 1 = 2,8708$$

Tra bảng 3 phần phụ lục $\chi_{0,05}^2$ với $df = (3- 1) (3-1)$ bậc tự do ($df = 4$) được giá trị $\chi_{(0,05;4)}^2 = 9,488$

$2,7807 < 9,488$ chấp nhận giả thiết H_0 "mật độ trồng cà phê khác nhau nhưng mức độ nhiễm bệnh gi sắt như nhau".

Trường hợp nếu có k công thức (k mẫu) nhưng số cặp $r = 2$ (số đặc trưng định tính là 2) A và $\bar{A}$ số liệu ghi như bảng (6.4)

$$\text{Thì: } \chi_m^2 = \frac{N^2}{Ta \times Tb} \left[\sum_{i=1}^k \frac{a_i^2}{n_i} - \frac{T_a^2}{N} \right] \quad (6.16)$$

$\chi_{t.n}^2$ được so sánh với χ^2 mức ý nghĩa α và bậc tự do $df = (k -1)$ tương ứng. Nếu $\chi_{t.n}^2 < \chi_{(\alpha,k-1)}^2$ chấp nhận giả thiết H_0 có nghĩa là hai đặc trưng định tính của các công thức là như nhau. Ngược lại $\chi_{t.n}^2 \geq \chi_{(\alpha,k-1)}^2$ bác bỏ giả thiết H_0 nghĩa là "các đặc trưng định tính của các cây trồng nghiên cứu là khác nhau" ở mức ý nghĩa.

Thí dụ: Trong nhà máy chế biến thực phẩm, người ta đã tiến hành điều tra chất lượng sản phẩm được sản xuất ra từ các dây chuyền khác nhau (1, 2, 3, 4) của cùng một doanh nghiệp chất lượng được phân làm 2 loại là thành phẩm (A) và phế phẩm ($B = \bar{A}$). Kết quả cụ thể như sau:

Bảng 6.4. Các tần số thực nghiệm

Dây chuyền \ Chỉ tiêu	1	2	3	4	T_i
A	75	82	65	82	304
B	15	12	10	9	46
n_i	90	94	75	91	350

Giả thiết H_0 đặt ra là: chất lượng sản phẩm của các dây chuyền sản xuất có thuần nhất không.

Giả thiết H_0 này được kiểm tra bằng tiêu chuẩn χ^2 như biểu thức (6.16)

$$\chi^2_m = \frac{N^2}{Ta \times Tb} \left[\sum_{i=1}^K \frac{a_i^2}{n_i} - \frac{T_a^2}{N} \right]$$

Thay số vào:

$$\chi^2_m = \frac{350^2}{304 \times 46} \left[\left(\frac{75^2}{90} + \frac{82^2}{94} + \frac{65^2}{75} + \frac{82^2}{91} \right) - \frac{304^2}{350} \right] = 1,804$$

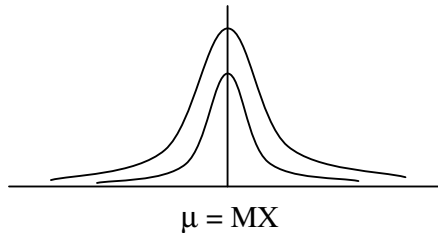
$\chi^2_{(\alpha, df)}$ ở mức ý nghĩa $\alpha = 0,05$ bậc tự do $df = 4-1 = 3$ là 7,815.

$\chi^2_{t,n} < \chi^2_{(0,05;3)}$ ($1.804 < 7,815$) nên ta chấp nhận giả thiết H_0 nghĩa là "các dây chuyền sản xuất khác nhau của doanh nghiệp có chất lượng sản phẩm được sản xuất ra là như nhau" ở mức 5% (hay $\alpha = 0,05$).

6.5. Kiểm tra giả thiết về hai phương sai

Hay còn gọi là so sánh 2 phương sai

Phương sai là một trong hai đặc trưng chính của tổng thể. Phương sai đại diện cho mức độ phân tán của các giá trị ngẫu nhiên của tổng thể nó biểu trưng một đặc điểm căn bản của các hiện tượng biến thiên nội tại trên một bản chất đồng nhất. Có nhiều trường hợp gặp hai đám đông có trung bình như nhau, nhưng phương sai của chúng lại khác nhau nên hai đám đông đó cũng khác nhau (hình 6.1).



Vì vậy, muốn xem xét 2 phương sai cần phải kiểm định để có thể đưa ra được nhận định hai phương sai ấy có đồng nhất (như nhau) hay không?

Thí dụ: Có hai tổng thể thực nghiệm (hai công thức) là công thức I gồm có n_1 cá thể có phương sai mẫu s_1^2 , công thức II có n_2 cá thể với phương sai mẫu s_2^2 . Vấn đề cần đặt ra là sự chênh lệch hay khác nhau giữa hai phương sai s_1^2 và s_2^2 là do một nguyên nhân nào đấy hay chỉ là sự khác nhau ngẫu nhiên (do rút mẫu ngẫu nhiên, còn 2 phương sai lý thuyết bằng nhau). Để giải quyết vấn đề này phải kiểm định giả thiết.

$$H_0: \sigma_1^2 = \sigma_2^2 \text{ hay } \frac{\sigma_1^2}{\sigma_2^2} = \frac{\sigma_2^1}{\sigma_1^2} = 1 \quad (6.17)$$

Đôi thiết $H_1: \sigma_1^2 \neq \sigma_2^2$ bằng kiểm định F (Fisher - Snedecor).

Phương sai s^2 là một đại lượng ngẫu nhiên có phân phối χ^2 với $n - 1$ bậc tự do. Như vậy, ta sẽ có 2 đại lượng s_1^2 và s_2^2 với 2 đại lượng χ^2 như sau:

Đại lượng χ_1^2 với bậc tự do $df_1 = n_1 - 1$

Đại lượng χ_2^2 với bậc tự do $df_2 = n_2 - 1$

Giả sử 2 phương sai lý thuyết bằng nhau $\sigma_1^2 = \sigma_2^2$. Thiết lập tỷ số giữa 2 phương sai mẫu là $F_{t,n} = \frac{s_1^2}{s_2^2}$. Tỷ số này cũng chính là tỷ số $\frac{\chi_1^2 \times df_1}{\chi_2^2 \times df_2} = \frac{\chi_2^2 \times df_2}{\chi_1^2 \times df_1}$ (6.18)

Tỷ số giữa 2 phương sai là đại lượng ngẫu nhiên tuân theo luật phân phối Fisher (F). Như vậy dựa vào bảng tính sẵn F (bảng 5) phần phụ lục có thể tìm được ngưỡng so sánh $F_{(\alpha; df_1, df_2)}$ ứng với mức ý nghĩa α và 2 bậc tự do df_1 và df_2 .

df_1 là bậc tự do của phương sai tử số (cột trong bảng F)

df_2 là bậc tự do của phương sai mẫu số (hàng trong bảng F)

Nếu $F_{t,n} < F_{lt}$ thì hai phương sai bằng nhau hay sự khác nhau giữa hai phương sai là ngẫu nhiên. Nếu $F_{t,n} \geq F_{lt}$ thì hai phương sai khác nhau (không bằng nhau).

(Cần lưu ý khi tính $F_{t,n}$ thì chọn phương sai tử số có giá trị số học lớn hơn phương sai mẫu số).

Thí dụ: Để tìm hiểu sự khác nhau giữa năng suất mủ cao su (năng suất cá thể: kg/cây) của hai điều kiện địa hình trồng khác nhau là đất bằng và đất dốc 5%, người ta tiến hành khảo sát trên đất bằng $n_1 = 29$ cây, có $s_1^2 = 0,35$ (kg/cây); trên đất dốc $n_2 = 21$ cây có $s_2^2 = 0,65$ (kg/cây). Hai điều kiện trồng cao su khác nhau thì phương sai của năng suất cá thể có bằng nhau với mức ý nghĩa $\alpha = 0,05$.

Xuất phát từ giả thiết H_0 là hai phương sai bằng nhau, như vậy áp dụng biểu thức (6.18) cụ thể là giá trị $F_{t,n}$

$$F_{t,n} = \frac{0,65}{0,35} = 1,857$$

Với bậc tự do $df_1 = n_2 - 1 = 20$ (cột 20) và bậc tự do $df_2 = 28$ (hàng 28) ta tìm được giá trị $F_{(0,05;20,28)} = 1,96$. Như vậy $F_{t,n} < F_{(0,05;20,28)}$ chúng ta chấp nhận giả thiết H_0 nghĩa là trồng cao su trên đất bằng và đất dốc 5% thì phương sai của năng suất cá thể (lượng mủ cao su cao) là bằng nhau.

CHƯƠNG VII

PHƯƠNG PHÁP SẮP XẾP CÔNG THỨC THÍ NGHIỆM VÀ PHÂN TÍCH KẾT QUẢ.

Mục đích của chương này là

Giúp cho sinh viên nắm vững phương pháp thiết kế một thí nghiệm nghiên cứu cơ bản, phương pháp bố trí thí nghiệm theo các kiểu khác nhau, biết cách phân tích phương sai (ANOVA) kết quả của thí nghiệm. Từ các kiến thức lý thuyết biết vận dụng vào thực tiễn việc bố trí thí nghiệm trong phòng cũng như trên đồng ruộng, biết cách phân tích kết quả thí nghiệm sau khi kết thúc và cách công bố kết quả nghiên cứu trên tạp chí khoa học hoặc trong các báo cáo khoa học.

Nội dung của chương:

1. Giới thiệu các kiểu thiết kế, sắp xếp thí nghiệm
 - Nhóm thí nghiệm một nhân tố:
 - i. Thiết kế kiểu hoàn toàn ngẫu nhiên (CRD)
 - ii. Thiết kế kiểu khối ngẫu nhiên đầy đủ (RCB)
 - iii. Thiết kế kiểu ô vuông la tinh (LS)
 - Nhóm thí nghiệm hai nhân tố:
 - i. Kiểu tổ hợp các mức của hai nhân tố có kiểu thiết kế hoàn toàn ngẫu nhiên (CRD) và khối ngẫu nhiên đầy đủ (RCB)
 - ii. Kiểu Chia ô (Split-Plot)
2. Phương pháp phân tích phương sai (ANOVA) tương ứng với các kiểu thiết kế
3. Cách so sánh các trung bình của các công thức và công bố kết quả phân tích, so sánh trên các báo cáo khoa học, luận văn tốt nghiệp hoặc tạp chí khoa học.

7.1. Các thí nghiệm một nhân tố.

7.1.1. Khái niệm:

Thí nghiệm một nhân tố là những thí nghiệm chỉ có một nhân tố thay đổi trong lúc tất cả các nhân tố khác được giữ nguyên. Như vậy, trong thí nghiệm này các công thức sẽ chứa các mức độ khác nhau của nhân tố đó, tất cả các nhân tố khác được áp dụng chung như nhau (ở cùng 1 mức bắt buộc) cho tất cả các công thức.

Ví dụ các thí nghiệm thử nghiệm giống cây trồng là thí nghiệm một nhân tố, ở đây giống là nhân tố thay đổi với các mức khác nhau (các giống khác nhau), có nghĩa là ở các công thức khác nhau, trên các mảnh thí nghiệm khác nhau, còn các nhân tố khác như quản lý, phân bón, phòng trừ sâu, bệnh... phải được áp dụng như nhau cho các mảnh thí nghiệm. Hay một ví dụ khác: nghiên cứu lượng phân đạm ảnh hưởng tới cây trồng thì một số lượng đạm thay đổi qua các công thức cho các mảnh thí nghiệm, còn lại tất cả các nhân tố khác được giữ cố định ở mức như nhau.

Hoặc một thí nghiệm so sánh mật độ cây trồng chẳng hạn thì mật độ thay đổi qua các công thức, còn các yếu tố khác được giữ cố định ở mức độ như nhau cho các mảnh thí nghiệm.

Đối với thí nghiệm một nhân tố, có thể thiết kế hoàn toàn ngẫu nhiên (CRD), thường dùng khi có ít công thức và khi có nhiều đơn vị thí nghiệm đồng nhất. Kiểu thiết kế thứ hai là

khối ngẫu nhiên bao gồm khối ngẫu nhiên đầy đủ (RCBD) và kiểu khối không đầy đủ (RIBD) sẽ được giới thiệu ở phần sau, kiểu thiết kế thứ ba là ô vuông la tinh (LS) sẽ được giới thiệu ở phần sau.

7.1.2. Các phương pháp sắp xếp và phân tích kết quả.

7.1.2.1. Thí nghiệm thiết kế kiểu hoàn toàn ngẫu nhiên (CRD)

Khái niệm

Một thí nghiệm được thiết kế kiểu hoàn toàn ngẫu nhiên khi các công thức được chỉ định một cách hoàn toàn ngẫu nhiên vào các ô thí nghiệm sao cho mỗi mảnh (ô) thí nghiệm đều có cơ hội như nhau để nhận được bất kỳ một công thức nào. Theo kiểu sắp xếp này, bất kỳ sự khác nhau nào (ngoài nhân tố thí nghiệm) giữa các ô thí nghiệm đều được coi là sai số thí nghiệm. Kiểu sắp xếp này chỉ phù hợp khi các đơn vị (ô, mảnh) thí nghiệm hoàn toàn đồng nhất (thường là thí nghiệm trong phòng), còn thí nghiệm ngoài đồng, thường có sự biến động lớn giữa các đơn vị thí nghiệm nên không hay áp dụng kiểu sắp xếp này.

Quá trình sắp xếp

Để minh họa qui trình sắp xếp và tiến tới vẽ sơ đồ của việc sắp xếp, ta dùng một ví dụ: thí nghiệm có 4 công thức A, B, C, D mỗi công thức được nhắc lại 5 lần, việc sắp xếp tiến hành theo các trình tự sau:

+ Bước 1. Xác định tổng số ô thí nghiệm cần có N.

$$N = r * t$$

Trong đó: N: tổng số ô thí nghiệm

t: số công thức cho mỗi lần nhắc lại

r: số lần nhắc lại của mỗi công thức

ở ví dụ này ta có: $N = r * t = 5 * 4 = 20$ ô như trong hình 7.1

1	2	3	4
5	6	7	8
9	10	11	12
13	14	15	16
17	18	19	20

Hình 7.1. Sơ đồ mẫu cho thí nghiệm thiết kế kiểu hoàn toàn ngẫu nhiên có 4 công và thức và 5 lần nhắc lại

Ấn định các công thức vào các ô thí nghiệm bằng một trong các cách ngẫu nhiên sau:

Cách 1: Lấy bảng số ngẫu nhiên (bảng 1 phụ lục)

- Từ điểm xuất phát bất kỳ trong bảng, đọc thẳng xuống dưới lấy 20 số ngẫu nhiên liên tục có 3 chữ số. (Ta lấy 3 chữ số để cho không có số nào trùng với số thứ tự từ 1 đến 20 là các số có 1 và 2 chữ số đã được ấn định cho các ô ở bước 2). Trong ví dụ này, 20 số ngẫu nhiên có 3 chữ số cùng thứ tự xuất hiện của chúng được ghi lại như bảng sau:

Thứ tự xuất hiện	Số ngẫu nhiên	Thứ tự xuất hiện	Số ngẫu nhiên
1	937	11	918
2	149	12	772
3	908	13	243
4	361	14	494
5	953	15	704
6	749	16	549
7	180	17	957
8	951	18	157
9	018	19	571
10	427	20	226

b. Xếp thứ tự theo hạng tăng dần hoặc giảm dần 20 số ngẫu nhiên ở bước 2. Trong ví dụ này xếp hạng thứ tự từ nhỏ nhất (số 1) đến lớn nhất (số 20) như sau:

Thứ tự xuất hiện	Số ngẫu nhiên	Xếp hạng	Thứ tự xuất hiện	Số ngẫu nhiên	Xếp hạng
1	937	17	11	918	16
2	149	2	12	772	14
3	908	15	13	243	6
4	361	7	14	494	9
5	953	19	15	704	12
6	749	13	16	549	10
7	180	4	17	957	20
8	951	18	18	157	3
9	018	1	19	571	11
10	427	8	20	226	5

c. Chia N số xếp hạng ở bước 3 thành t nhóm, mỗi nhóm chứa r số thứ tự mà các số ngẫu nhiên xuất hiện. Trong ví dụ này, 20 số xếp hạng được chia thành 4 nhóm, mỗi nhóm chứa 5 số như sau:

Nhóm	Thứ tự xếp hạng				
1	17	2	15	7	19
2	13	4	18	1	8
3	16	14	6	9	12
4	10	20	3	11	5

d. Cho mỗi nhóm mang tên một công thức, nhóm 1 mang tên công thức A, nhóm 2 mang tên công thức B, nhóm 3 mang tên công thức C và nhóm 4 mang tên công thức D. Thứ tự xếp hạng được coi như thứ tự của các ô trong sơ đồ hình 7.1. Như vậy, kết quả là công thức A

nằm ở các vị trí ô số 17, 2, 15, 7 và 19. Công thức B nằm ở các vị trí ô số 13, 4, 18, 1, và 8. Cứ tiếp tục như vậy để chỉ định vị trí của các công thức còn lại.

Cách 2. Lấy một cỗ bài

a. Rút ra N con bài, mỗi lần rút một con sau đó trộn lại cỗ bài và rút con khác (không bỏ con đã rút vào). Phương pháp này áp dụng khi $N < 52$. Trong ví dụ này 20 con được rút ra cùng thứ tự của chúng như sau.

Thứ tự xuất hiện	1	2	3	4	5	6	7	8	9	10
Tên các con bài	J ♥	3 ♦	A ♦	R ♥	Q ♣	5 ♥	6 ♣	9 ♠	9 ♥	8 ♠

Thứ tự xuất hiện	11	12	13	14	15	16	17	18	19	20
Tên các con bài	A ♥	4 ♦	3 ♥	2 ♣	7 ♣	10 ♠	A ♣	4 ♠	6 ♥	10 ♣

b. Xếp hạng 20 con bài rút ra từ bước 1 theo thứ tự số từ 2 đến Át, nếu cùng số thì theo thứ tự của chất bài..

Trong ví dụ này, 20 con được xếp hạng tương ứng với thứ tự xuất hiện như sau:

Thứ tự xuất hiện và xếp hạng

1	2	3	4	5	6	7	8	9	10
14	7	9	15	5	11	2	19	13	18

Thứ tự xuất hiện và xếp hạng

11	12	13	14	15	16	17	18	19	20
16	8	10	1	3	20	6	17	12	4

c. Chia các số xếp hạng thành 4 nhóm, gán công thức và xếp các công thức vào các ô theo như cách 1.

Trong ví dụ này được kết quả thu được như sau:

Công thức	Án định vào các ô số				
A	14	7	9	15	5
B	11	2	19	13	18
C	16	8	10	1	3
D	20	6	17	12	4

Cách 3. Rút thăm.

a. Chuẩn bị N mẫu giấy, chia mẫu giấy thành t nhóm, các mẫu giấy trong mỗi nhóm có cùng ký hiệu của một công thức. Trong ví dụ này sẽ có 5 mẫu mang chữ A, 5 mẫu chữ B, ... Trộn lẫn 20 mẫu giấy trong một hộp (các mẫu giấy được gấp kín).

b. Rút mỗi lần một mẫu giấy, đặt vào các ô theo trật tự từ đầu đến cuối. Mở mảnh giấy ra, ta có công thức được chỉ định vào các ô như sau:

Thứ tự xuất hiện (ô) công thức theo các số như sau:

1	2	3	4	5	6	7	8	9	10
D	B	A	B	C	A	D	C	B	D
11	12	13	14	15	16	17	18	19	20
D	A	A	B	B	C	D	C	C	A

Trong ví dụ này như công thức A nằm ở các ô 3, 6, 12, 13, và 20.

Phân tích phương sai

Có 2 nguồn biến động trong N quan sát thu được từ thí nghiệm sắp xếp kiểu CRD. Một nguồn là biến động do công thức, nguồn thứ 2 là sai số thí nghiệm (do các yếu tố ngẫu nhiên tác động). Độ lớn của 2 nguồn biến động này được so sánh để biết liệu sự khác nhau của các quan sát từ các công thức có thực sự hay chỉ là sự khác nhau ngẫu nhiên. Sự khác nhau của các công thức là thực sự nếu biến động của các công thức đủ lớn hơn biến động của sai số thí nghiệm.

Sắp xếp thí nghiệm kiểu này có thuận lợi là đơn giản cho quá trình phân tích phương sai, đặc biệt là khi số lần nhắc lại của các công thức không bằng nhau.

1) Số lần nhắc lại bằng nhau.

Các bước phân tích phương sai cho thí nghiệm sắp xếp kiểu CRD khi số lần nhắc lại bằng nhau được minh họa bằng thí nghiệm dùng thuốc hoá học trừ rầy nâu và sâu đục thân lúa, với 4 lần nhắc lại và 7 công thức (6 loại thuốc mới và 1 đối chứng), số liệu được ghi trong bảng 7.1. Bảng 7.1. Năng suất lúa ở các công thức và các lần nhắc lại khác nhau.

Thứ tự công thức	Năng xuất hạt (kg/ ha)				Tổng công thức(T_i)	Trung bình công thức($\bar{x}_i$)
	1	2	3	4		
T_1	2.537	2.069	2.104	1.797	8.507	2.127
T_2	3.366	2.591	2.211	2.544	10.712	2.678
T_3	2.536	2.459	2.827	2.385	10.207	2.552
T_4	2.387	2.453	1.556	2.116	8.512	2.128
T_5	1.997	1.679	1.649	1.859	7.184	1.796
T_6	1.796	1.704	1.904	1.320	6.724	1.681
T_7	1.401	1.516	1.270	1.077	5.264	1.316
Tổng toàn bộ (G)					57.110	2.040
Trung bình toàn bộ						

Ghi chú: T₁: Dol- Mix (1kg)

T₂: Dol- Mix (2kg)

T₃: DDT+γ - BHC

T₄: Azodrin

T₅:Dimecron- Boom

T₆: Dimecron- knap

T₇: Đồi chứng

r: Số nhắc lại = 4

t: số công thức= 7

Các bước tính toán như sau:

Bước 1. Tính các tổng T_i của các công thức; tổng toàn bộ G và các $\bar{x}_i$ (trung bình của các công thức) như trong bảng 7.1.

Bước 2. Đưa ra một bảng phân tích phương sai như sau:

Nguồn biến động	Bậc tự do (df)	Tổng bình phương (SS)	Bình phương Trung bình (MS)	F _{tn}	F _{bảng}	
					5%	1%
Công thức (T)						
Sai số (E)						
Toàn bộ (To)						

Bước 3. Tính các bậc tự do (df).

- Bậc tự do của toàn bộ dfTo = N-1 = r * t - 1 = 4 * 7 - 1 = 27
- Bậc tự do của công thức dfT = t - 1 = 7 - 1 = 6
- Bậc tự do của sai số dfE = t (r - 1) = 7(4 - 1) = 21 hoặc bằng bậc tự do toàn bộ - bậc tự do công thức = 27 - 6 = 21

Bước 4. x_{ij} là số liệu của công thức i, lần lặp j

T_i: tổng các x_{ij} của công thức thứ i

N là tổng số mảnh thí nghiệm: N = t * r = 7 * 4 = 28

Tính số điều chỉnh (CF) và các loại tổng bình phương (SS) như sau:

$$CF = \frac{G^2}{N} = 116484004$$

$$SSTo = \sum_{i=1}^r x_{ij}^2 - CF = (2537^2 + 2069^2 + \dots + 1270^2) - 116484004 = 7577412$$

$$SST = \sum_{i=1}^t \frac{T_i^2}{r} - CF = \frac{8507^2 + 10712^2 + \dots + 5264^2}{4} - 116484004 = 5587174$$

$$SSE = SSTo - SST = 1990238$$

Bước 5. Tính các bình phương trung bình (MS) cho mỗi nguồn biến động.

$$MST = \frac{SST}{t-1} = 931196$$

$$MSE = \frac{SSE}{t(r-1)} = 94773$$