Bảng 7.11: Phát thảo phân tích phương sai cho thí nghiệm thừa số 2 x 4 x 3 theo kiểu bố trí lô phụ trong lô phụ

Nguồn biến động	Độ tự do	Tổng bình phương	Trung bình bình phương	F(tính)	F(bảng) 5%	F(bảng) 1%
Phân tích lô chính:						
Lặp lại	r – 1 = 2					
Nhân tố lô chính (A)	a – 1 = 2					
Sai số (a)	(r – 1)(a – 1) = 4					
Phân tích lô phụ:						
Nhân tố lô phụ (B)	(b – 1) = 1					
A x B	(a – 1)(b - 1) = 2					
Sai số (b)	a(r – 1)(b – 1) = 6					
Phân tích lô phụ trong lô phụ:						
Nhân tố lô phụ trong lô phụ (C)	(c – 1) = 2					
A x C	(a – 1)(c - 1) = 4					
B x C	(b – 1)(c - 1) = 2					
A x B x C	(a – 1)(b – 1)(c – 1) = 4					
Sai số (c)	ab(r – 1)(c – 1) = 24					
Tổng cộng	rabc – 1 = 53					

Bảng 7.12: Tổng năng suất của lặp lại x thời điểm cấy (RA) được tính từ số liệu ở bảng 7.10.

Thời điểm cấy	Tổng năng suất (RA) Lặp lại I	Lặp lại II	Lặp lại III	Tổng TĐ cấy (A)
P_1	199,9	206,6	191,0	597,5
P_2	223,9	193,2	189,9	607,0
P_3	164,9	174,6	155,7	495,2
Tổng lặp lại (R)	588,7	574,4	536,6	1699,7

$$\text{SS tổng cộng} = \sum X^2 - \text{C.F.}$$

$$= [(25,7)^2 + \cdots + (34,8)^2] - 53499,631 = 2048,559$$

$$\text{SS lặp lại} = \frac{\sum R^2}{abc} - \text{C.F.}$$

$$= \frac{(588,7)^2 + (574,4)^2 + (536,6)^2}{(3)(2)(3)} - 53499,631 = 80,514$$

$$\text{SS A (thời điểm cấy)} = \frac{\sum A^2}{rbc} - \text{C.F.}$$

$$= \frac{(597,5)^2 + (607,0)^2 + (495,2)^2}{(3)(2)(3)} - 53499,631 = 426,94$$

$$\text{SS sai số (a)} = \frac{\sum (RA)^2}{bc} - \text{C.F.} - \text{SS lặp lại} - \text{SS A}$$

$$= \frac{(199,9)^2 + \cdots + (155,7)^2}{6} - 53499,631 - 80,514 - 426,94 = 86,863$$

Bước 2: Thực hiện phân tích lô phụ.

- Lập bảng tổng thời điểm cấy x biện pháp kiểm soát rầy (AB) xem Bảng 7.13. Tính tổng biện pháp kiểm soát rầy (B).

Bảng 7.13: Tổng năng suất của thời điểm cấy x biện pháp kiểm soát rầy (AB) được tính từ số liệu ở bảng 7.10.

Biện pháp kiểm soát (B)	Tổng năng suất (AB) P_1	P_2	P_3	**Tổng (B)**
S_0	265,8	283,5	229,5	**778,8**
S_1	331,7	323,5	265,7	**920,9**

- Lập bảng tổng lặp lại x thời điểm cấy x biện pháp kiểm soát rầy (RAB), Bảng 7.14.

Bảng 7.14: Tổng năng suất của lặp lại x thời điểm cấy x biện pháp kiểm soát rầy (RAB) được tính từ số liệu ở bảng 7.10.

Thời điểm cấy	Tổng năng suất (RAB) Lặp lại I	Lặp lại II	Lặp lại III
		S_0	
P_1	92,1	92,1	81,6
P_2	104,8	88,7	90,0
P_3	78,5	81,8	69,2
		S_1	
P_1	107,8	114,5	109,4
P_2	119,1	104,5	99,9
P_3	86,4	92,8	86,5

- Tính các tổng bình phương.

$$\text{SS B(kiểm soát rầy)} = \frac{\sum B^2}{rac} - \text{C.F.}$$

$$= \frac{(778,8)^2 + (920.9)^2}{(3)(3)(3)} - 53499,631 = 373,933$$

$$\text{SS AxB} = \frac{\sum (AB)^2}{rc} - \text{C.F.} - \text{SS(A)} - \text{SS(B)}$$

$$= \frac{(265,8)^2 + \cdots + (265,7)^2}{(3)(3)} - 53499,631 - 426,94 - 373,933 = 29,025$$

$$\text{SS sai số (b)} = \frac{\sum (RAB)^2}{c} - \text{C.F.} - \text{SS lặp lại} - \text{SS A} - \text{SS sai số (a)} - \text{SS B} - \text{SS AxB}$$

$$= \frac{(92,1)^2 + \cdots + (86,5)^2}{3} - 53499,631 - 80,514 - 426,940 - 86,863 - 373,933 - 29,025$$

$$= 23,03$$

Bước 4: Lập 3 bảng tổng

- Bảng tổng thời điểm cấy x thời điểm thu hoạch (A x C). Tính tổng thời điểm thu hoạch (C), Bảng 7.15.
- Bảng tổng biện pháp kiểm soát rầy x thời điểm thu hoạch (B x C), Bảng 7.16.
- Bảng tổng thời điểm cấy x biện pháp kiểm soát rầy x thời điểm thu hoạch (AxBxC), Bảng 7.17.

Bảng 7.15: Tổng thời điểm cấy x thời điểm thu hoạch (AC) được tính từ số liệu ở bảng 7.10.

Thời điểm cấy (A)	Tổng năng suất (AC)		
	H_1	H_2	H_3
P_1	163,1	203,2	231,2
P_2	179,9	200,3	226,8
P_3	137,8	164,2	193,2
Tổng C	480,8	567,7	651,2

Bảng 7.16: Tổng năng suất biện pháp kiểm soát rầy x thời điểm thu hoạch (BC) được tính từ số liệu ở bảng 7.10

Biện pháp kiểm soát rầy	Tổng năng suất (BC)		
	H_1	H_2	H_3
S_0	225,1	266,7	287,0
S_1	255,7	301,0	364,2

Bảng 7.17: Tổng năng suất của thời điểm cấy x biện pháp kiểm soát rầy x thời điểm thu hoạch (ABC) được tính từ số liệu ở bảng 7.10.

Thời điểm cấy (A)	Tổng năng suất (ABC)		
	H_1	H_2	H_3
		S_0	
P_1	74,9	90,0	100,9
P_2	81,4	100,0	102,1
P_3	68,8	76,7	84,0
		S_1	
P_1	88,2	113,2	130,3
P_2	98,5	100,3	124,7
P_3	69,0	87,5	109,2

$$\text{SS C (thời điểm thu hoạch)} = \frac{\sum C^2}{rab} - C.F.$$

$$= \frac{(480,8)^2 + (567,7)^2 + (651,2)^2}{(3)(3)(2)} - 53499,631 = 806,667$$

$$\text{SS AxC} = \frac{\sum (AC)^2}{rb} - C.F. - SS(A) - SS(C)$$

$$= \frac{(163,1)^2 + \cdots + (193,2)^2}{(3)(2)} - 53499,631 - 426,94 - 806,667 = 24,153$$

$$\text{SS BxC} = \frac{\sum (BC)^2}{ra} - C.F. - SS(B) - SS(C)$$

$$= \frac{(225,1)^2 + \cdots + (364,2)^2}{(3)(3)} - 53499,631 - 373,933 - 806,667 = 74,549$$

$$\text{SS AxBxC} = \frac{\sum (ABC)^2}{r} - C.F. - SS(A) - SS(B) - SS(C)$$

$$= \frac{(74,9)^2 + \cdots + (109,2)^2}{(3)} - 53499,631 - 426,94 - 373,933 - 806,667 = 44,904$$

SS sai số(c) = SS tổng cộng - (tất cả SS đã tính)

= 2048,559 - (80,514 + 426,94 + 86,863 + 373,933 + 29,025 + 23,03 + 806,667 + 24,153 + 74,549 + 44,904) = 77,98

Bước 5: Đối với mỗi nguồn biến động, tính các giá trị trung bình bình phương bằng cách tổng bình phương (SS) cho độ tự do tương ứng (df).

Bước 7: Tính trung bình bình phương cho từng nguồn biến động bằng cách chia mỗi tổng bình phương cho độ tự do tương ứng.

$$\text{MS lặp lại} = \frac{\text{SS lặp lại}}{r-1} = \frac{80,514}{2} = 40,257$$

$$\text{MS}(A) = \frac{\text{SS}(A)}{a-1} = \frac{426,94}{2} = 213,47$$

$$\text{MS sai số (a)} = \frac{\text{SS sai số(a)}}{(r-1)(a-1)} = \frac{86,863}{4} = 21,716$$

$$\text{MS}(B) = \frac{\text{SS}(B)}{b-1} = \frac{373,933}{1} = 373,933$$

$$\text{MS AxB)} = \frac{\text{SS (AxB)}}{(a-1)(b-1)} = \frac{29,025}{2} = 14,512$$

$$\text{MS sai số (b)} = \frac{\text{SS sai số(b)}}{(r-1)(b-1)} = \frac{29,025}{2} = 14,512$$

$$\text{MS}(C) = \frac{\text{SS}(C)}{c-1} = \frac{806,227}{2} = 403,333$$

$$\text{MS}(AxC) = \frac{\text{SS}(AxC)}{(a-1)(c-1)} = \frac{24,153}{4} = 6,038$$

$$\text{MS AxBxC)} = \frac{\text{SS}(AxBxC)}{(a-1)(b-1)(c-1)} = \frac{74,549}{2} = 37,275$$

$$\textbf{MS sai số (c)} = \frac{\textbf{SS sai số (c)}}{ab(r-1)(c-1)} = \frac{77,98}{24} = 3,249$$

Bước 5: Tính các giá trị F như sau:

$$F(A) = \frac{\text{MS}(A)}{\text{MS sai số (a)}} = \frac{213,47}{21,716} = 9,83$$

$$F(B) = \frac{\text{MS}(B)}{\text{MS sai số (b)}} = \frac{373,933}{3,838} = 97,42$$

$$F(AxB) = \frac{\text{MS}(AxB)}{\text{MS sai số (b)}} = \frac{14,512}{3,838} = 3,78$$

$$F(C) = \frac{\text{MS}(C)}{\text{MS sai số (c)}} = \frac{403,333}{3,249} = 124,13$$

$$F(AxC) = \frac{\text{MS}(AxC)}{\text{MS sai số (c)}} = \frac{6,038}{3,249} = 1,86$$

$$\text{MS}(BxC) = \frac{\text{MS}(BxC)}{\text{MS sai số (c)}} = \frac{37,275}{3,249} = 11,47$$

$$F(AxBxC) = \frac{\text{MS}(AxBxC)}{\text{MS sai số (c)}} = \frac{11,226}{3,249} = 3,46$$

Bước 7: So sánh trị số F tính của mỗi ảnh hưởng với trị số F bảng tương ứng, trong Phụ lục VI, với df_1 là độ tự do của MS tử số và df_2 là độ tự do của MS mẫu số, ở mức ý nghĩa α . Ví dụ, trị số F bảng tương ứng với F(A) có $df_1 = 2$ và $df_2 = 4$ là 6,94 ở $\alpha = 0,05$ và 18,00 ở $\alpha = 0,01$ (Bảng 7.18).

Bước 10: Tính ba hệ số biến động tương ứng với ba dạng sai số.

$$cv(a) = \frac{\sqrt{MS\ sai\ số\ (a)}}{\bar{x}} x\,100 = \frac{\sqrt{21,716}}{31,48} x100 = 14,8\%$$

$$cv(b) = \frac{\sqrt{MS\ sai\ số\ (b)}}{\bar{x}} x\,100 = \frac{\sqrt{3,838}}{31,48} x100 = 6,2\%$$

$$cv(c) = \frac{\sqrt{MS\ sai\ số\ (c)}}{\bar{x}} x\,100 = \frac{\sqrt{3,249}}{31,48} x100 = 5,7\%$$

Bảng 7.18: Phân tích phương sai cho thí nghiệm thừa số 3 x 2 x 3 theo kiểu bố trí lô phụ trong lô phụ với ba lần lặp lại.

Nguồn biến động	Độ tự do	Tổng bình phương	Trung bình bình phương	F(tính)	F(bảng) 5%	1%
Phân tích lô chính:						
Lặp lại	2	80,514	40,257	$1,85^{ns}$	6,94	18,00
Nhân tố lô chính (A)	2	426,940	213,470	9,83*	6,94	18,00
Sai số (a)	4	86,863	21,716			
Phân tích lô phụ:						
Nhân tố lô phụ (B)	1	373,933	373,933	97,42**	5,99	13,74
A x B	2	29,025	14,512	$3,78^{ns}$	5,14	10,92
Sai số (b)	6	23,030	3,838			
Phân tích lô phụ trong lô phụ:						
Nhân tố lô phụ trong lô phụ (C)	2	806,667	403,333	124,13**	3,40	5,61
A x C	4	24,153	6,038	$1,86^{ns}$	2,78	4,22
B x C	2	74,549	37,275	11,47**	3,40	5,61
A x B x C	4	44,904	11,226	3,46*	2,78	4,22
Sai số (c)	24	77,980	3,249			
Tổng cộng	53	2048,559				

cv(a) = 14,8%, cv(b) = 6,2% và cv(c) = 5,7%

* = Ý nghĩa ở mức 5%, ** = Ý nghĩa ở mức 1% và ^{ns} = không ý nghĩa

CHƯƠNG 8

SO SÁNH CÁC TRUNG BÌNH NGHIỆM THỨC

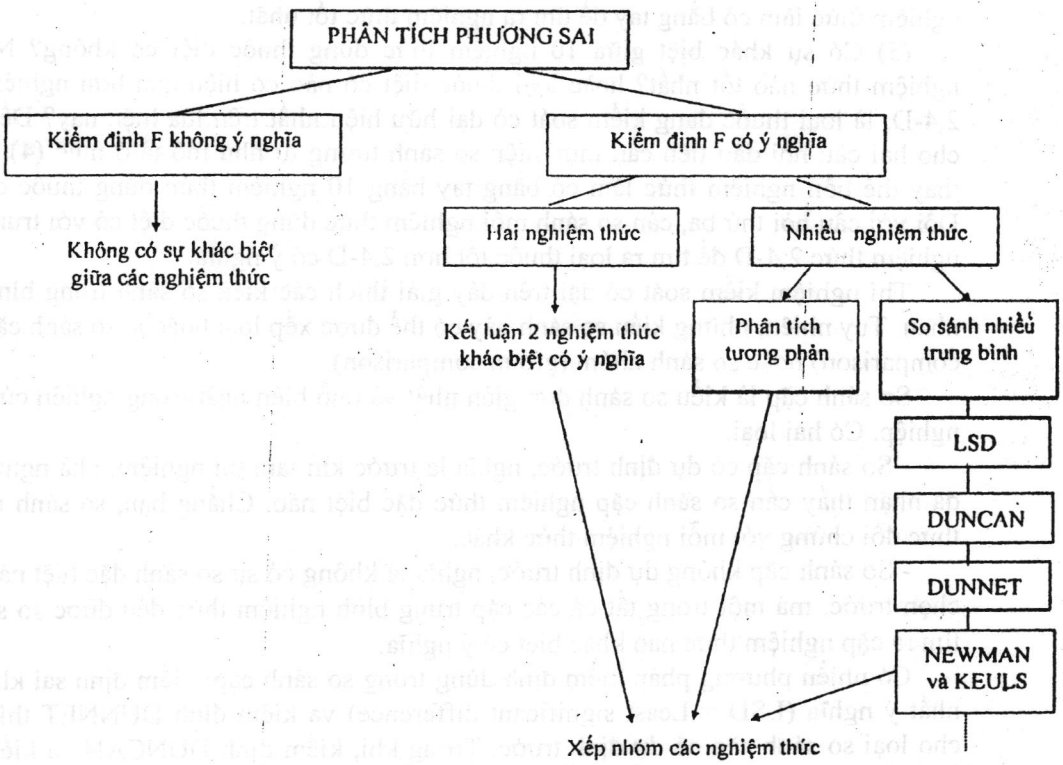
Kiểm định F trong phân tích phương sai chỉ cho kết luận tổng quát: loại hoặc không loại giả thuyết không (H_0: Các trung bình bằng nhau)

- Nếu kiểm định F không ý nghĩa (F tính < F bảng), chúng ta kết luận các nghiệm thức khác biệt không ý nghĩa (ở mức α đã chọn) và phân tích được kết thúc.
- Nếu kiểm định F có ý nghĩa, phân tích phương sai chưa cho kết luận đầy đủ.

Tuy nhiên, nếu chỉ có hai trung bình nghiệm thức, kiểm định F có ý nghĩa đủ để kết luận hai trung bình khác biệt có ý nghĩa (ở mức α đã chọn) và phân tích cũng chấm dứt.

Trái lại, nếu có nhiều trung bình nghiệm thức, phân tích phương sai chưa cho kết luận hoàn toàn. Vì thế, cần phân tích tiếp để biết các cặp nghiệm thức nào khác biệt có ý nghĩa.

Phương pháp phân tích có thể được tóm tắt như sau:



Như vậy, khi kiểm định F trong phân tích phương sai có ý nghĩa và có nhiều hơn hai trung bình nghiệm thức, cần phân tích tiếp để đi đến kết luận cụ thể cho từng cặp trung bình nghiệm thức.

Nhìn chung, có nhiều kiểu so sánh các trung bình nghiệm thức, song cần xác định kiểu so sánh nào cần thiết để có thể trả lời được mục tiêu của thí nghiệm. Ví dụ, thí nghiệm kiểm soát cỏ dại trên lúa với 15 nghiệm thức: bốn nghiệm thức nhổ cỏ bằng

tay, 10 nghiệm thức dùng thuốc trừ cỏ và một nghiệm thức không làm cỏ. Các câu hỏi có thể được đặt ra là:

(1) Có nghiệm thức nào kiểm soát cỏ dại hiệu quả không? Để trả lời câu hỏi này, chúng ta có thể thực hiện so sánh trung bình của nghiệm thức không làm cỏ với trung bình của từng nghiệm thức còn lại.

(2) Có sự khác biệt nào giữa 14 nghiệm thức có kiểm soát cỏ dại không? Nếu có, nghiệm thức nào hữu hiệu và nghiệm thức nào không hữu hiệu? Trong số các nghiệm thức hữu hiệu, có sự khác biệt ý nghĩa giữa chúng không? Nếu có, nghiệm thức nào tốt nhất? Các câu hỏi trên được trả lời bằng cách thực hiện so sánh mỗi trung bình của nghiệm thức có kiểm soát cỏ dại với trung bình của nghiệm thức đối chứng (không làm cỏ) và chọn các nghiệm thức kiểm soát cỏ dại tốt hơn đối chứng. Mặt khác, so sánh các nghiệm thức đã chọn để tìm ra nghiệm thức kiểm soát cỏ dại tốt nhất.

(3) Có sự khác biệt giữa nhóm nghiệm thức làm cỏ bằng tay và nhóm nghiệm thức dùng thuốc diệt cỏ không? Để có thể kết luận được vấn đề này cần so sánh trung bình của bốn trung bình nghiệm thức làm cỏ bằng tay với trung bình của 10 trung bình nghiệm thức dùng thuốc diệt cỏ.

(4) Có sự khác biệt giữa bốn nghiệm thức làm cỏ bằng tay không? Nếu có, nghiệm thức nào tốt nhất? Để trả lời các câu hỏi như thế, cần so sánh bốn trung bình của nghiệm thức làm cỏ bằng tay để tìm ra nghiệm thức tốt nhất.

(5) Có sự khác biệt giữa 10 nghiệm thức dùng thuốc diệt cỏ không? Nếu có, nghiệm thức nào tốt nhất? hoặc loại thuốc diệt cỏ nào có hiệu quả hơn nghiệm thức 2,4-D, là loại thuốc đang kiểm soát cỏ dại hữu hiệu nhất trên lúa hiện nay? Để trả lời cho hai câu hỏi đầu tiên cần thực hiện so sánh tương tự như mô tả ở mục (4), nhưng thay thế bốn nghiệm thức làm cỏ bằng tay bằng 10 nghiệm thức dùng thuốc diệt cỏ. Đối với câu hỏi thứ ba, cần so sánh mỗi nghiệm thức dùng thuốc diệt cỏ với trung bình nghiệm thức 2,4-D để tìm ra loại thuốc tốt hơn 2,4-D có ý nghĩa.

Thí nghiệm kiểm soát cỏ dại trên đây giải thích các kiểu so sánh trung bình khác nhau. Tuy nhiên, những kiểu so sánh này có thể được xếp loại hoặc là so sánh cặp (pair comparison) hoặc so sánh nhóm (group comparison).

So sánh cặp là kiểu so sánh đơn giản nhất và phổ biến nhất trong nghiên cứu nông nghiệp. Có hai loại:

- So sánh cặp có dự định trước, nghĩa là trước khi làm thí nghiệm, nhà nghiên cứu đã nhận thấy cần so sánh cặp nghiệm thức đặc biệt nào. Chẳng hạn, so sánh nghiệm thức đối chứng với mỗi nghiệm thức khác.
- So sánh cặp không dự định trước, nghĩa là không có sự so sánh đặc biệt nào được chọn trước, mà một trong tất cả các cặp trung bình nghiệm thức đều được so sánh để tìm ra cặp nghiệm thức nào khác biệt có ý nghĩa.

Có nhiều phương pháp kiểm định dùng trong so sánh cặp: kiểm định sai khác nhỏ nhất ý nghĩa (LSD = Least significant difference) và kiểm định DUNNET thích hợp cho loại so sánh cặp có dự định trước. Trong khi, kiểm định DUNCAN và kiểm định NEWMAN và KEULS,... có thể áp dụng cho kiểu so sánh cặp không dự định trước.

I. KIỂM ĐỊNH KHÁC BIỆT NHỎ NHẤT CÓ Ý NGHĨA

(Least Significant Difference Test = LSD)

Kiểm định LSD là phương pháp được dùng phổ biến nhất và đơn giản nhất để thực hiện việc so sánh cặp. Phương pháp này cho một giá trị LSD duy nhất ở mức ý nghĩa

α, nó được dùng như ranh giới giữa sự khác biệt có ý nghĩa và không ý nghĩa của bất kỳ cặp trung bình nghiệm thức nào. Nghĩa là, hai nghiệm thức được xem là khác biệt có ý nghĩa ở mức α, nếu sự sai khác của chúng vượt quá giá trị LSD tính; trái lại, chúng sẽ khác biệt không ý nghĩa.

Ví dụ: Dùng phương pháp kiểm định LSD để so sánh hai trung bình nghiệm thức i và j, các bước thực hiện như sau:

Bước 1: Tính sai biệt giữa hai trung bình nghiệm thức i và j

$$d_{ij} = \overline{X_i} - \overline{X_j}$$

Bước 2: Tính giá trị LSD ở mức ý nghĩa α

$$LSD_\alpha = (t_{\alpha/2})(s_{\bar{d}})$$

với $t_{\alpha/2}$ = giá trị t bảng (2 đuôi) với độ tự do của trung bình bình phương sai số ở mức ý nghĩa α

$s_{\bar{d}}$ = sai số chuẩn của sai biệt hai trung bình

Bước 3: So sánh sự sai biệt của hai trung bình tính được ở bước 1 với giá trị LSD tính ở bước 2. Nếu trị tuyệt đối của d_{ij} lớn hơn giá trị LSD, ta kết luận hai nghiệm thức i và j khác biệt ý nghĩa; trái lại, sẽ khác biệt không ý nghĩa.

Sai số chuẩn của sai biệt hai trung bình ($s_{\bar{d}}$) sẽ thay đổi theo kiểu bố trí thí nghiệm và số lần lặp lại của hai nghiệm thức được so sánh.

1. Thí nghiệm một nhân tố *(CRD, RCBD và hình vuông latin)*

a) Số lần lặp lại bằng nhau

Đối với các kiểu bố trí này chỉ có một loại sai số, nên giá trị $s_{\bar{d}}$ của bất kỳ cặp trung bình nghiệm thức nào cũng được tính như sau:

$$s_{\bar{d}} = \sqrt{\frac{2s_e^2}{r}}$$

với s_e^2 là trung bình bình phương sai số và r là số lần lặp lại.

Ví dụ: Số liệu của thí nghiệm bố trí ngẫu nhiên đầy đủ với bảy nghiệm thức (sáu nghiệm thức thuốc trừ sâu và một nghiệm thức đối chứng), bốn lần lặp lại (chương 5, Bảng 5.1). Giả sử, mục đích chính của thí nghiệm là xác định nghiệm thức thuốc trừ sâu nào tốt hơn nghiệm thức đối chứng; nghĩa là, so sánh một trong sáu trung bình nghiệm thức thuốc trừ sâu với trung bình của nghiệm thức đối chứng.

Bước 1: Tính sai biệt trung bình giữa nghiệm thức đối chứng với một trong sáu nghiệm thức thuốc trừ sâu, như ở Bảng 8.1

Bước 2: Tính trị số LSD ở mức ý nghĩa α

$$LSD_\alpha = t_{\alpha/2}\sqrt{\frac{2s_e^2}{r}}$$

Theo ví dụ, trung bình bình phương sai số là 0,095; độ tự do sai số bằng 21 và 4 lần lặp lại. Xem Phụ lục IV, trị số t bảng với độ tự do 21 là 2,080 ở mức α = 0,05 và 2,831 ở mức α = 0,01. Các giá trị LSD được tính như sau:

$$LSD_{0,05} = 2,080\sqrt{\frac{2(0,095)}{4}} = 0,453 \text{ t/ha}$$

$$LSD_{0.01} = 2,831\sqrt{\frac{2(0,095)}{4}} = 0,617 \text{ t/ha}$$

Bảng 8.1: So sánh năng suất trung bình của nghiệm thức đối chứng với một trong sáu nghiệm thức thuốc trừ sâu, dùng kiểm định LSD (số liệu ở Bảng 5.1).

Nghiệm thức	Năng suất trung bình tấn/ha (a)	Khác biệt với đối chứng tấn/ha
Dol-Mix (1kg)	2,127	0, 811**
Dol-Mix (2kg)	2,678	1,362**
DDT + -BHC	2,552	1,236**
Azodrin	2,128	0,812**
Dimecron-Boom	1,796	0,480*
Dimecron-Knap	1,681	0,365^{ns}
Đối chứng	1,316	-

(a) Trung bình của 4 lần lặp lại ** = khác biệt ở mức ý nghĩa 1%
* = khác biệt ở mức ý nghĩa 5% ^{ns} = khác biệt không ý nghĩa

Bước 3: So sánh mỗi giá trị sai khác trung bình tính ở bước 1 với giá trị LSD tính ở bước 2, và đánh dấu **, * hoặc ns trên trị số sai khác, với:

** : Khi giá trị sai khác trung bình lớn hơn giá trị LSD ở mức ý nghĩa 1%.

* : Khi giá trị sai khác trung bình lớn hơn giá trị LSD ở mức ý nghĩa 5%.

^{ns} : Khi giá trị sai khác trung bình nhỏ hơn giá trị LSD ở mức ý nghĩa 5%.

Trong ví dụ, kết quả của sáu so sánh cặp cho thấy ngoại trừ Dimecron-Knap, tất cả các nghiệm thức thuốc trừ sâu còn lại đều cho năng suất cao hơn nghiệm thức đối chứng có ý nghĩa (Bảng 8.1).

Mặc dầu, kiểm định LSD thích hợp để thực hiện so sánh cặp có dự định trước, nghĩa là so sánh bất kỳ nghiệm thức nào với nghiệm thức đối chứng. Nhưng đôi khi nó cũng được dùng để so sánh tất cả các cặp trung bình nghiệm thức; đặc biệt, khi số nghiệm thức ít. Khi số nghiệm thức nhiều, số cặp trung bình nghiệm thức sẽ gia tăng nhanh theo số nghiệm thức. Ví dụ, với 5 nghiệm thức sẽ có 10 cặp trung bình nghiệm thức, với 10 nghiệm thức sẽ có 45 cặp trung bình nghiệm thức, và với 15 nghiệm thức sẽ có 105 cặp trung bình nghiệm thức; do đó, nếu dùng kiểm định LSD sẽ không có giá trị.

Chú ý, nếu phải sử dụng LSD để so sánh tất cả các cặp trung bình nghiệm thức, chỉ áp dụng khi kiểm định F của nghiệm thức có ý nghĩa và số nghiệm thức không quá lớn (thường $t \leq 5$).

Ví dụ: Sử dụng số liệu của bố trí khối ngẫu nhiên đầy đủ (RCBD), bốn lần lặp lại với năm nghiệm thức là năm cách bón phân đạm khác nhau. Các bước thức hiện như sau:

Bước 1: Xếp các trung bình nghiệm thức theo thứ tự tăng dần (hoặc giảm dần). Ví dụ, năm trung bình nghiệm thức được xếp theo thứ tự tăng dần như sau:

Nghiệm thức	Năng suất trung bình tạ/ha	Xếp hạng
E	38,0	1
B	49,8	2
A	49,9	3
D	51,5	4
C	52,6	5

Bước 2: Tính giá trị LSD ở mức ý nghĩa α đã định

$$LSD_{\alpha} = t_{\alpha/2}\sqrt{\frac{2s_e^2}{r}}$$

Theo ví dụ, trung bình bình phương sai số là 2,837; độ tự do sai số bằng 12, và 4 lần lặp lại. Xem Phụ lục IV, trị số t bảng với độ tự do 12 là 2,179 ở mức α = 0,05. Giá trị LSD được tính như sau:

$$LSD_{0,05} = 2,179\sqrt{\frac{2(2,837)}{4}} = 2,595 \text{ t/ha}$$

Bước 3: Tính sai biệt của tất cả các cặp trung bình có thể có bằng cách lập bảng trung bình đã được xếp thứ tự, với cột bên trái bỏ trị số lớn nhất và hàng trên cùng bỏ trị số nhỏ nhất. Trừ từng trung bình ở cột bên trái với các trung bình của hàng trên cùng và ghi các sai biệt vào trong bảng (xem Bảng 8.2).

Bảng 8.2: Các sai biệt của 5 trung bình nghiệm thức đã được xếp hạng.

	49,8	49,9	51,5	52,6
38,0	11,8*	11,9*	13,5*	14,6*
49,8		$0,1^{ns}$	$1,7^{ns}$	2,8*
49,9			$1,6^{ns}$	2,7*
51,5				$1,1^{ns}$

So sánh từng giá trị sai biệt này với các giá trị LSD ở mức α = 0,05 và đánh dấu (*) nếu giá trị sai khác lớn hơn giá trị LSD và ns nếu sai khác nhỏ hơn giá trị LSD (Bảng 8.2).

Bước 5: Trình bày kết quả kiểm định theo một trong hai cách sau đây:

(1) Dùng ký hiệu đoạn thẳng nếu trình bày kết quả theo thứ tự xếp hạng bằng cách trong từng hàng, dùng đoạn thẳng nối tất cả trung bình mà qua kiểm định LSD khác biệt không ý nghĩa.

38,0 49,8 49,9 51,5 52,6

Vì đường thẳng thứ hai nằm trong đường thứ nhất nên bỏ đường thứ hai. Kết quả sau cùng được trình bày như sau:

38,0 49,8 49,9 51,5 52,6

(2) Dùng ký hiệu alphabet nếu trình bày kết quả không theo thứ tự xếp hạng. Để đơn giản, mỗi đoạn thẳng hoặc mỗi trung bình đứng độc lập được ghi một chữ. Thường chúng ta dùng chữ a cho giá trị trung bình cao nhất, kế tiếp là chữ b,... Ví dụ, ghi a, b, c cho hai đoạn thẳng và một trung bình độc lập như sau:

38,0 49,8 49,9 51,5 52,6
.c b a

Sau cùng, trình bày kết quả theo ký hiệu alphabet như ở Bảng 8.3. Chú ý, mỗi nghiệm thức có thể được ghi một hay nhiều chữ. Ví dụ, một chữ (a) cho nghiệm thức C, và hai chữ (ab) cho nghiệm thức D.

Bảng 8.3: Dùng kiểm định LSD để so sánh tất cả các cặp trung bình nghiệm thức. Số liệu trong Bảng 5.1 của thí nghiệm RCRD gồm 5 nghiệm thức (sử dụng ký hiệu alphabet)

Nghiệm thức	Năng suất trung bình kg/ha	LSD
E	38,0	c
B	49,8	b
A	49,9	b
D	51,5	ab
C	52,6	a

Các trung bình có cùng chữ theo sau thì khác biệt không ý nghĩa ở mức $\alpha = 0.05$

b) Khi hai nghiệm thức so sánh có số lần lặp lại không bằng nhau như trong kiểu bố trí hoàn toàn ngẫu nhiên (CRD), LSD sẽ được tính khác nhau cho mỗi cặp trung bình nghiệm thức.

Ví dụ: Sử dụng số liệu của bố trí hoàn toàn ngẫu nhiên với năm nghiệm thức có số lần lặp lại không bằng nhau (chương 5, Bảng 5.3). Giả sử chúng ta muốn so sánh bất kỳ cặp trung bình nghiệm thức nào, LSD sẽ được tính theo hai cách sau đây:

* Để so sánh hai trung bình nghiệm thức có số lần lặp lại bằng nhau, LSD được tính giống như trên.

$$LSD_\alpha = t_{\alpha/2}\sqrt{\frac{2s_e^2}{r}}$$

Chẳng hạn, so sánh trung bình của nghiệm thức 1 với nghiệm thức 5 có 3 lần lặp lại, giá trị LSD_α được tính như sau:

$$LSD_{0,05} = 2,179\sqrt{\frac{2(6,451)}{3}} = 4,519(kg/ha)$$

* Để so sánh hai trung bình nghiệm thức có số lần lặp lại không bằng nhau, LSD được tính như sau:

$$LSD_\alpha = (t_{\alpha/2})(s_{\bar{d}})$$

với:

$$s_{\bar{d}} = \sqrt{s_e^2\left(\frac{1}{r_i} + \frac{1}{r_j}\right)}$$

với: r_i và r_j lần lượt là số lần lặp lại của nghiệm thức thứ i và thứ j

Ví dụ: So sánh trung bình của nghiệm thức 1 (3 lần lặp lại) với nghiệm thức 2 (4 lần lặp lại):

$$LSD_{0,05} = 2,179\sqrt{6,451\left(\frac{1}{3}+\frac{1}{4}\right)} = 4,227 \text{ kg/ha}$$

Tính tương tự cho các cặp trung bình nghiệm thức khác.

Ví dụ: Để theo dõi tất cả sự khác biệt có thể có giữa các trung bình nghiệm thức, chúng ta xếp thứ tự các số trung bình từ nhỏ đến lớn và lập bảng sai biệt của các cặp trung bình như Bảng 8.2 như sau:

	40,00	41,00	48,00	48,33
31,75	8,25	9,25	16,25	16,58
40,00		1,00	8,00	8,33
41,00			7,00	7,33
48,00				0,33

So sánh từng giá trị sai biệt này với các giá trị LSD ở mức ý nghĩa 5% và đánh dấu (*) nếu giá trị sai khác lớn hơn giá trị LSD và ns nếu sai khác nhỏ hơn giá trị LSD. Bảng sau cùng có dạng như sau:

	40,00	41,00	48,00	48,33
31,75	8,25*	9,25*	16,25*	16,58*
40,00		$1,00^{ns}$	8,00*	8,33*
41,00			7,00*	7,33*
48,00				$0,33^{ns}$

Dùng ký hiệu đoạn thẳng để nối các trung bình khác biệt không ý nghĩa theo từng hàng sai biệt:

31,75 40,00 41,00 48,00 48,33

Dùng ký hiệu alphabet cho mỗi đoạn thẳng và trung bình đứng độc lập. Kết quả sau cùng được trình bày trong Bảng 8.4.

31,75 40,00 41,00 48,00 48,33

c ——— b ——— a

Bảng 8.4: Kiểm định LSD để so sánh tất cả các cặp trung bình nghiệm thức. Thí nghiệm điều tra về sự tăng trọng của thỏ ở năm địa điểm khác nhau. Dùng ký hiệu alphabet (Số liệu ở Bảng 5.3)

Địa điểm	Trọng lượng trung bình (oz. /con)
1	41,00 b
2	31,75 c
3	48,00 a
4	40,00 b
5	48,33 a

Các trung bình có cùng chữ theo sau thì khác biệt không ý nghĩa ở mức $\alpha = 0.05$

c) Cách tính $s_{\bar{d}}$ trong trường hợp có số liệu thiếu

- Bố trí khối hoàn toàn ngẫu nhiên (RCBD)

* Trường hợp có một số liệu bị thiếu, cách tính sai số chuẩn của sai biệt hai trung bình ($s_{\bar{d}}$)

$$s_{\bar{d}} = \sqrt{s_e^2[\frac{2}{r} + \frac{t}{r(r-1)(t-1)}]}$$

Ví dụ: Để so sánh trung bình của nghiệm thức 3 (nghiệm thức có chứa số liệu thiếu) với một trong các nghiệm thức còn lại (Chương 5, Bảng 5.8), $s_{\bar{d}}$ được tính như sau:

$$s_{\bar{d}} = \sqrt{2{,}524\left[\frac{2}{4} + \frac{5}{(4)(3)(4)}\right]} = 1{,}235 \text{ kg/ha}$$

Để tính trị số LSD, chúng ta dùng giá trị t bảng ở mức ý nghĩa α, với độ tự do bằng 11. Trị số LSD để so sánh nghiệm thức 4 với bất kỳ nghiệm thức nào khác được tính như sau:

$LSD_{\alpha} = (t_{\alpha/2})(s_{\bar{d}})$

$LSD_{0,05} = (2{,}201)(1{,}235) = 2{,}718$ kg/ha

$LSD_{0,01} = (3{,}106)(1{,}235) = 3{,}836$ kg/ha

* Trường hợp so sánh cặp mà cả hai nghiệm thức đều có một số liệu thiếu, hoặc một trong hai nghiệm thức có hai số liệu thiếu thì $s_{\bar{d}}$ được tính như sau:

$$s_{\bar{d}} = \sqrt{s_e^2(\frac{1}{r_A'} + \frac{1}{r_B'})}$$

với r_A' và r_B' lần lượt là số lần lặp lại hữu hiệu của nghiệm thức A và B. Đối với nghiệm thức được so sánh, chẳng hạn nghiệm thức A, lặp lại được tính bằng 1 nếu nó có số liệu ở cả hai nghiệm thức, bằng 1/2 nếu nó có số liệu ở nghiệm thức A nhưng không có số liệu ở nghiệm thức B và bằng 0 nếu không có số liệu ở nghiệm thức A.

Ví dụ: Nếu nghiệm thức A được so sánh với nghiệm thức C (nghĩa là, cả hai nghiệm thức đều có một số liệu thiếu), cách tính như sau:

(1) Định số lần lặp lại hữu hiệu cho mỗi nghiệm thức. Ví dụ, với nghiệm thức A, lần lặp lại thứ nhất được tính bằng 0 (vì nghiệm thức A có số liệu thiếu ở lần lặp lại thứ nhất), lần lặp lại thứ hai được tính bằng 1/2 (vì nghiệm thức A không có số liệu thiếu nhưng nghiệm thức C có số liệu thiếu) và lần lặp lại thứ ba và thứ tư được tính bằng 1 (vì cả hai nghiệm thức đều không có số liệu thiếu ở cả hai lần lặp lại này). Do đó, số lần lặp lại hữu hiệu của nghiệm thức A sẽ bằng 0 + 1/2 + 1 + 1 = 2,5. Tính tương tự số lần lặp lại hữu hiệu cho nghiệm thức C cũng bằng 2,5.

(2) Tính $s_{\bar{d}}$:

$$s_{\bar{d}} = \sqrt{s_e^2(\frac{1}{2{,}5} + \frac{1}{2{,}5})} = \sqrt{(2{,}699)(0{,}8)} = 2{,}159$$

- Bố trí hình vuông Latin

$$s_{\bar{d}} = \sqrt{s_e^2[\frac{2}{t} + \frac{1}{(t-1)(t-2)}]}$$

Ví dụ: Để so sánh nghiệm thức A (nghiệm thức có số liệu thiếu) với một trong các nghiệm thức còn lại $s_{\bar{d}}$, được tính như sau:

$$s_{\bar{d}} = \sqrt{0,025\left[\frac{2}{4} + \frac{1}{(4-1)(4-2)}\right]} = 0,13$$

2. Thí nghiệm thừa số hai nhân tố

Đối với thí nghiệm thừa số có nhiều loại trung bình nghiệm thức cần so sánh.

a. Bố trí khối hoàn toàn ngẫu nhiên (CRD và RCBD)

Với thí nghiệm thừa số a x b (nhân tố A có a mức độ và nhân tố B có b mức độ). Có ba loại trung bình có thể được so sánh: Các trị số LSD của ba loại so sánh này như sau:

(1) So sánh a trung bình của a mức độ ở nhân tố A (trung bình được tính trên b mức độ của nhân tố B và r lần lặp lại).

$$LSD_{\alpha}(A) = t_{\alpha/2}\sqrt{\frac{2s_e^2}{rb}}$$

(2) So sánh b trung bình của b mức độ ở nhân tố B (trung bình được tính trên a mức độ của nhân tố A và r lần lặp lại).

$$LSD_{\alpha}(B) = t_{\alpha/2}\sqrt{\frac{2s_e^2}{ra}}$$

(3) So sánh (a x b) trung bình (trung bình được tính trên r lần lặp lại); trong đó, so sánh a trung bình của nhân tố A ở mỗi mức độ của nhân tố B, hoặc so sánh b trung bình của nhân tố B ở mỗi mức độ của nhân tố A.

$$LSD_{\alpha}(AxB) = t_{\alpha/2}\sqrt{\frac{2s_e^2}{r}}$$

với $t_{\alpha/2}$ = giá trị t bảng (2 đuôi) với độ tự do của trung bình bình phương sai số ở mức ý nghĩa α

* Trường hợp có ảnh hưởng tương tác, chỉ cần tính ảnh hưởng đơn của một nhân tố ở từng mức độ của nhân tố thứ hai.

Ví dụ: Thí nghiệm so sánh năng suất của 2 giống lúa được bón ở 3 mức phân đạm (Thí nghiệm thừa số 2 x 3) với 3 lần lặp lại ở Bảng 6.3, chương 6. Kết quả phân tích phương sai cho thấy tương tác có ý nghĩa, chứng tỏ ảnh hưởng của giống thay đổi theo mức độ đạm được áp dụng. Vì vậy, so sánh trung bình giữa các mức độ đạm được tính dựa trên trung bình của tất cả các giống, hoặc trung bình giữa các giống được tính dựa trên trung bình của tất cả các mức độ đạm sẽ không hiệu quả, mà loại so sánh thích hợp hơn là so sánh giữa các trung bình giống ở cùng mức độ đạm hoặc giữa các trung bình mức độ đạm của cùng giống. Chẳng hạn, chúng ta muốn so sánh sự khác biệt giữa các mức phân đạm cho từng giống; lúc đó giá trị LSD được tính như sau:

$$LSD_{\alpha}(AxB) = t_{\alpha/2}\sqrt{\frac{2s_e^2}{r}}$$

$$LSD_{0,05} = 2,179\sqrt{\frac{(2)(0,127)}{(3)}} = 0,634$$

Lập bảng trung bình giống x đạm, dùng trị số LSD để so sánh các trung bình của đạm ở cùng giống như trong Bảng 8.5.

Bảng 8.5: Năng suất trung bình của 2 giống lúa được trắc nghiệm với 3 mức độ đạm của bố trí hoàn toàn ngẫu nhiên với 3 lần lặp lại (Bảng 6.3)

Giống	Năng suất trung bình (AxB)[a]			
	N_0	N_1	N_2	TB. giống (A)
V_1	6,38 b	10,25 a	10,67 a	9,10
V_2	7,53 c	9,30 b	10,03 a	8,95
TB, đạm (B)	6,96	9,77	10,35	

Các trung bình trong cùng một hàng có cùng chữ theo sau thì khác biệt không ý nghĩa qua kiểm định LSD ở mức $\alpha = 0.05$

[a] Trung bình của 3 lần lặp lại

* Trường hợp không có ảnh hưởng tương tác, chỉ cần tính các ảnh hưởng chính của nhân tố A hoặc của nhân tố B.

Ví dụ: So sánh năng suất của ba giống lúa được trắc nhiệm ở năm mức độ đạm" (thừa số 3 x 5) với bốn lần lặp lại ở Bảng 6.7, chương 6. Kết quả phân tích phương sai cho thấy tương tác không có ý nghĩa; vì vậy, chỉ cần so sánh trung bình giữa các mức độ đạm được tính dựa trên trung bình của tất cả các giống, hoặc trung bình giữa các giống được tính dựa trên trung bình của tất cả các mức độ đạm. Các giá trị LSD tính được như sau:

(1) So sánh 3 trung bình của 3 giống

$$LSD_{0,05} = 2,0189\sqrt{\frac{(2)(0,151)}{(4)(5)}} = 0,248$$

(2) So sánh 5 trung bình của 5 mức độ đạm

$$LSD_{0,05} = 2,0189\sqrt{\frac{(2)(0,151)}{(4)(3)}} = 0,320$$

Lập bảng trung bình giống x đạm, dùng các trị số LSD để so sánh các trung bình của giống và các trung bình của đạm như trong Bảng 8.6. Các bước kiểm định LSD được thực hiện tương tự như trong thí nghiệm một nhân tố. Kết quả kiểm định LSD cho nhân tố giống và đạm được trình bày trong Bảng 8.6.

Bảng 8.6: Năng suất trung bình của 3 giống lúa được trắc nghiệm ở 5 mức độ đạm. Bố trí RCBD với 4 lần lặp lại (Số liệu ở Bảng 6.7)

Đạm	Trung bình năng suất (AxB)			
	V_1	V_2	V_3	TB đạm (B)[a]
N_0	3,124	3,548	3,778	3,483 d
N_1	4,724	4,806	4,754	4,761 c
N_2	4,460	5,436	5,320	5,072 b
N_3	5,717	5,740	5,553	5,670 a
N_4	5,823	5,681	5,887	5,797 a
TB giống (A)[b]	4,769 b	5,042 a	5,058 a	

Các trung bình trong cùng một hàng hoặc trong cùng một cột có cùng chữ theo sau thì khác biệt không ý nghĩa qua kiểm định LSD ở mức $\alpha = 0.05$

[a] Trung bình của 3 giống và 4 lần lặp lại

[b] Trung bình của 5 mức độ đạm và 4 lần lặp lại

b. Bố trí lô phụ

Trong bố trí lô phụ, có hai nhân tố và hai dạng sai số nên có bốn loại so sánh cặp khác nhau. Mỗi loại đòi hỏi tính một trị số LSD riêng. Bảng 8.7 trình bày công thức tính LSD = $ts_{\bar{d}}$ của bốn loại so sánh này.

Bảng 8.7: Công thức tính LSD của bốn loại so sánh cặp ở bố trí lô phụ.

Số T.T.	Loại so sánh cặp giữa	LSD = $t.s_{\bar{d}}$
1	2 trung bình lô chính (tính trung bình trên tất cả nghiệm thức lô phụ): $a_1 - a_2$	$t_a\sqrt{\frac{2E_a}{rb}}$
2	2 trung bình lô phụ (tính trung bình tính trên tất cả các nghiệm thức lô chính): $b_1 - b_2$	$t_b\sqrt{\frac{2E_b}{ra}}$
3	2 trung bình lô phụ ở cùng nghiệm thức lô chính ($b_1a_1 - b_2a_1$)	$t_b\sqrt{\frac{2E_b}{r}}$
4	2 trung bình lô chính ở cùng hoặc khác nghiệm thức lô phụ. Hay 2 trung bình lô phụ ở các nghiệm thức lô chính khác nhau ($b_1a_1 - b_1a_2$ hoặc $b_1a_1 - b_2a_2$)	$t'\sqrt{\frac{2[(b-1)E_b + E_a}{rb}}$

* E_a = MS sai số (a), E_b = MS sai số (b), t_a = t bảng ở độ tự do sai số (a), t_b = t bảng ở độ tự do sai số (b), r = số lần lặp lại, a = số nghiệm thức lô chính, b = số nghiệm thức lô phụ. Đối với so sánh loại (4), giá trị t' bảng có gia trọng được tính theo công thức sau:

$$t' = \frac{(b-1)E_b t_b + E_a t_a}{(b-1)E_b + E_a}$$

Ví dụ: Quan sát số liệu của thí nghiệm thừa số 6 x 4 trong Bảng 6.6 đến Bảng 6.8 của chương 6. Kết quả phân tích phương sai (Bảng 6.9) cho thấy tương tác giữa đạm và giống có ý nghĩa cao, chứng tỏ ảnh hưởng của giống thay đổi theo mức độ đạm được áp dụng. Vì vậy, loại so sánh thích hợp hơn là so sánh giữa các trung bình giống ở cùng mức độ đạm hoặc giữa các trung bình mức độ đạm của cùng giống. Tuy nhiên, vì so sánh cặp giữa các trung bình mức độ đạm của cùng giống không thích hợp do đặc tính định lượng của các nghiệm thức mức độ đạm; do đó, chỉ giải thích so sánh giữa các trung bình giống với cùng mức độ đạm.

Tính LSD để so sánh hai trung bình giống ở cùng mức độ đạm (nghĩa là, hai trung bình lô phụ ở cùng nghiệm thức lô chính)

Tính sai số chuẩn của sai khác hai trung bình ($s_{\bar{d}}$) theo công thức so sánh loại (3) ở Bảng 8.7 như sau:

$$s_{\bar{d}} = \sqrt{2E_b / r}$$

$$s_{\bar{d}} = \sqrt{(2)(349.580)/3} = 482{,}8(kg/ha)$$

Trị số t_b bảng, với độ tự do bằng 36, ở $\alpha = 0{,}05$ là 2,029 và ở $\alpha = 0{,}01$ là 2,722

$$LSD_{0,05} = (2{,}029)(482{,}8) = 980(kg/ha)$$

$$LSD_{0,01} = (2{,}722)(482{,}8) = 1.314(kg/ha)$$

Lập bảng trung bình giống x đạm, dùng các trị số LSD để so sánh hai trung bình giống ở cùng mức độ đạm như trong Bảng 8.8.

Bảng 8.8: Năng suất trung bình của 4 giống lúa được trắc nghiệm với 6 mức độ đạm trong bố trí lô phụ (Số liệu ở Bảng 6.10)

Mức độ đạm	Năng suất trung bình, kg/ha [a]			
kg/ha	IR8	IR5	C4-63	Peta
0	4,253 a	4,306 a	3,183 b	4,481 a
60	5,672 ab	5,982 a	5,443 ab	4,816 b
90	6,400 a	6,259 a	5,994 a	4,812 b
120	6,733 a	6,895 a	6,014 a	3,816 b
150	7,563 a	6,951 a	6,687 a	2,047 b
180	8,701 a	6,540 b	6,065 b	1,881 c

Các trung bình trong cùng một hàng có cùng chữ theo sau thì khác biệt không ý nghĩa qua kiểm định LSD ở mức $\alpha = 0.05$

[a] Trung bình của 3 lần lặp lại.

So sánh mỗi cặp giống (với cùng mức độ đạm), tính sai khác trung bình và so sánh giá trị này với giá trị LSD. Chẳng hạn, giữa giống Peta và IR8 ở 0 kgN/ha, sai khác trung bình là 4,481 - 4,253 = 228 kg/ha. Vì giá trị sai khác này nhỏ hơn giá trị LSD ở mức ý nghĩa 5%, nên hai trung bình này khác biệt không ý nghĩa.

Trường hợp có số liệu thiếu.

Ví dụ: Để so sánh trung bình của N_2 với trung bình của bất cứ mức độ đạm nào khác ở V_1, $s_{\bar{d}}$ được tính như sau:

$$s_{\bar{d}} = \sqrt{\frac{2\left\{E_e + E_b\left[(b-1) + \frac{b^2}{2(r-1)(b-1)}\right]\right\}}{rb}}$$

$$s_{\bar{d}} = \sqrt{\frac{2\left\{141.148 + 358.779\left[(4-1) + \frac{16}{2(2)(3)}\right]\right\}}{(3)(4)}} = 531{,}64 \text{ kg/ha}$$

Mặt khác, để so sánh trung bình của V_1 với trung bình của bất kỳ giống nào khác ở cùng mức độ đạm, $s_{\bar{d}}$ được tính như sau:

$$s_{\bar{d}} = \sqrt{\frac{2E_b\left[1 + \frac{b}{2a(r-1)(b-1)}\right]}{r}}$$

$$s_{\bar{d}} = \sqrt{\frac{2.(358.779)\left[1 + \frac{4}{(2)(6)(2)(3)}\right]}{3}} = 502{,}47 \text{ kg/ha}$$

c. Bố trí dải phụ

Tương tự như bố trí lô phụ, bố trí dải phụ có bốn loại so sánh cặp. Công thức tính giá trị $s_{\bar{d}}$ tương ứng cho mỗi loại được trình bày trong Bảng 8.9.

Phương pháp áp dụng kiểm định LSD để so sánh cặp ở bố trí dãy phụ được giải thích với thí nghiệm thừa số 6 x 3. Số liệu và kết quả phân tích phương sai được trình bày trọng Bảng 6.11 - 6.16 (Chương 6). Vì tương tác giữa giống và đạm có ý nghĩa (Bảng 6.16) nên chỉ có dạng so sánh cặp thích hợp là so sánh giữa các giống ở cùng mức độ đạm.

Các bước tính giá trị LSD để so sánh hai giống ở cùng mức độ đạm là:

Bước 1: Tính giá trị $s_{\bar{d}}$ theo công thức loại 3 của Bảng 8.9:

$$s_{\bar{d}} = \sqrt{\frac{2[(b-1)E_c + E_a}{rb}} = \sqrt{\frac{2[(2)(0,412)+1,492]}{(3)(3)}} = 0,717$$

Bảng 8.9: Công thức tính LSD của một trong bốn loại so sánh cặp trong bố trí dải phụ.

Số T.T.	Loại so sánh cặp giữa	LSD = t.$s_{\bar{d}}$
1	2 trung bình nhân tố ngang (được tính trung bình trên tất cả nghiệm thức nhân tố dọc)	$t_a\sqrt{\frac{2E_a}{rb}}$
2	2 trung bình nhân tố dọc (được tính trung bình trên tất cả nghiệm thức nhân tố ngang)	$t_b\sqrt{\frac{2E_b}{ra}}$
3	2 trung bình nhân tố ngang ở cùng mức độ của nhân tố dọc	$t'\sqrt{\frac{2[(b-1)E_c + E_a}{rb}}$
4	2 trung bình nhân tố dọc ở cùng mức độ của nhân tố ngang	$t^*\sqrt{\frac{2[(a-1)E_c + E_b}{ra}}$

E_a = MS sai số (a), E_b = MS sai số (b), E_c = MS sai số (c), t_a = t bảng ở độ tự do sai số (a), t_b = t bảng ở độ tự do sai số (b), t_c = t bảng ở độ tự do sai số (c), r = số lần lặp lại, a = số mức độ của nhân tố dải ngang, b = số mức độ của nhân tố dải dọc.

Đối với so sánh loại (3) và (4), giá trị t' và t* bảng có gia trọng được tính theo các công thức sau:

$$t' = \frac{(b-1)E_c t_c + E_a t_a}{(b-1)E_c + E_a} \quad \text{(Loại 3)} \qquad t^* = \frac{(a-1)E_c t_c + E_b t_b}{(a-1)E_c + E_b} \quad \text{(Loại 4)}$$

Bước 2: Vì trong công thức sử dụng ở bước 1 có hai dạng sai số (E_a và E_c), nên tính giá trị t bảng có gia trọng như sau:

*** Từ Phụ lục IV, tìm được giá trị t bảng tương ứng với E_a có n = 10 độ tự do (nghĩa là t_a) và giá trị t bảng tương ứng với E_c có n = 20 độ tự do (nghĩa là t_c) ở mức ý** nghĩa 5% và 1%:

$$t_a(0,05) = 2,228 \text{ và } t_a(0,01) = 3,169$$
$$t_c(0,05) = 2,086 \text{ và } t_c(0,01) = 2,845$$

* Tính giá trị t có gia trọng theo công thức tính ở lô phụ

$$t' = \frac{(b-1)E_c t_c + E_a t_a}{(b-1)E_c + E_a}$$

$$t'_{0,05} = \frac{(2)(0,412)(2,086)+(1,492)(2,228)}{(2)(0,412)+1,492} = 2,178$$

$$t'_{0,01} = \frac{(2)(0,412)(2,845)+(1,492)(3,169)}{(2)(0,412)+1,492} = 3,054$$

Bước 3: Tính giá trị LSD ở mức ý nghĩa 5% và 1%:

$$LSD_{0,05} = (t'_{0,05})(s_{\bar{d}}) = (2,178)(0,717) = 1,562$$
$$LSD_{0,01} = (t'_{0,01})(s_{\bar{d}}) = (3,054)(0,717) = 2,190$$

Bước 4: Lập bảng trung bình hai chiều giống x đạm (Bảng 8.10), với giá trị LSD được tính ở bước 3. Trong ví dụ, để xác định năng suất trung bình của giống IR8 có khác biệt ý nghĩa với giống Peta ở mức độ 0 kgN/ha hay không. Sự sai biệt trung bình của chúng (3,572 – 3,207 = 0,365 t/ha) được so với giá trị LSD 1,562 và 2,190 t/ha. Vì sự sai biệt trung bình nhỏ hơn giá trị LSD ở mức ý nghĩa 5%, nên có thể kết luận năng suất trung bình của IR8 và Peta ở 0 kgN/ha khác biệt không ý nghĩa.

Bảng 8.10: Năng suất trung bình của sáu giống lúa với ba mức độ đạm ở bố trí dải phụ (Số liệu trong Bảng 6.11)

	Năng suất trung bình[a] (t/ha)		
Giống	0 kgN/ha	60 kgN/ha	120 kgN/ha
IR 8	3,572	5,132	7,548
IR 127-80	4,934	6,714	7,211
IR 305-4 -12	4,250	6,122	7,868
IR 400-2-5	4,059	5,554	7,094
IR 665-58	4,102	5,633	6,012
Peta	3,207	3,714	2,492

[a] *Trung bình của 3 lần lặp lại. Giá trị LSD dùng để so sánh hai giống ở cùng mức độ đạm là 1,562 t/ha ở mức ý nghĩa 5% và 2,190 t/ha ở mức ý nghĩa 1%.*

b. Bố trí lô phụ trong lô phụ

Đối với bố trí lô phụ trong lô phụ, có 12 loại so sánh cặp, mỗi loại đòi hỏi tính một trị số LSD riêng. Các công thức tính $s_{\bar{d}}$ của từng loại so sánh được trình bày trong Bảng 8.11.

II. KIỂM ĐỊNH DUNNET

Kiểm định DUNNET là phương pháp so sánh t trung bình nghiệm thức với một đối chứng.

Phương pháp kiểm định LSD "Dunnet" (LSD_{DN}) được thực hiện tương tự như kiểm định LSD, ngoại trừ thay thế giá trị t bảng bằng giá trị d_t trong bảng Dunnet (Phụ lục IX).

$$LSD_{DN} = d_t \cdot \sqrt{\frac{2.s_e^2}{r}}$$

với: s_e^2 là phương sai sai số

r là số lần lặp lại của mỗi trung bình nghiệm thức

d_t xem trong bảng Dunnet (Phụ lục IX) ở độ tự do của phương sai sai số; trong đó, α là mức ý nghĩa được chọn và t là số trung bình nghiệm thức (không kể đối chứng) so với đối chứng.

Ví dụ: Sử dụng lại số liệu của thí nghiệm so sánh 6 nghiệm thức thuốc trừ sâu với một nghiệm thức đối chứng (Chương 5, Bảng 5.1). Dùng kiểm định Dunnet để so sánh sáu trung bình nghiệm thức này với đối chứng.

Kết quả phân tích phương sai cho trung bình bình phương sai số, s_e^2 = 94773, với độ tự do sai số bằng 21.

Vì có sáu nghiệm thức so với đối chứng, nên t = 6. Giá trị d_t trong bảng Dunnet ở α = 0,05 là 2,4525

Bảng 8.11: Sai số chuẩn của sai khác trung bình của một trong 12 loại so sánh cặp trong bố trí lô phụ trong lô phụ.

Số TT	Loại so sánh cặp giữa	$s_{\bar{d}}$
1	2 trung bình lô chính (trung bình trên tất cả nghiệm thức lô phụ và lô phụ trong lô phụ)	$\sqrt{\frac{2E_a}{rbc}}$
2	2 trung bình lô phụ (trung bình trên tất cả nghiệm thức lô chính và lô phụ-trong lô phụ)	$\sqrt{\frac{2E_b}{rac}}$
3	2 trung bình lô phụ (trung bình trên tất cả nghiệm thức lô phụ-trong lô phụ) ở cùng hoặc khác mức độ của nhạn tố lô chính	$\sqrt{\frac{2E_b}{rc}}$
4	2 trung bình lô phụ-trong lô phụ (trung bình trên tất cả nghiệm thức lô chính và lô phụ)	$\sqrt{\frac{2E_C}{rab}}$
5	2 trung bình lô phụ-trong lô phụ ở cùng mức độ của nhân tố lô chính (trung bình trên tất cả nghiệm thức lô phụ)	$\sqrt{\frac{2E_C}{rb}}$
6	2 trung bình lô phụ-trong lô phụ ở cùng mức độ của nhân tố lô phụ (trung bình trên tất cả nghiệm thức lô chính)	$\sqrt{\frac{2E_C}{ra}}$
7	2 trung bình lô phụ-trong lô phụ ở cùng phối hợp của các nghiệm thức lô chính và lô phụ.	$\sqrt{\frac{2E_C}{r}}$
8	2 trung bình lô chính (trung bình trên tất cả nghiệm thức lô phụ-trong lô phụ) ở cùng hoặc khác mức độ của nhân tố lô phụ	$\sqrt{\frac{2[(b-1)E_b+E_a}{rbc}}$
9	2 trung bình lô phụ (trung bình trên tất cả nghiệm thức lô chính) ở cùng hoặc khác mức độ của nhân tố lô phụ-trong lô phụ	$\sqrt{\frac{2[(c-1)E_c+E_b}{rac}}$
10	2 trung bình lô phụ ở cùng phối hợp của nghiệm thức lô chính và lô phụ-trong lô phụ	$\sqrt{\frac{2[(c-1)E_c+E_b}{rc}}$
11	2 trung bình lô chính (trung bình trên tất cả nghiệm thức lô phụ) ở cùng hoặc khác mức độ của nhân tố lô phụ-trong lô phụ	$\sqrt{\frac{2[(c-1)E_c+E_a}{rbc}}$
12	2 trung bình lô chính ở cùng phối hợp của nghiệm thức lô phụ và lô phụ-trong lô phụ	$\sqrt{\frac{2[b(c-1)E_c+(b-1)E_b+E_a}{rbc}}$

* E_a = MS sai số (a), E_b = MS sai số (b), E_c = MS sai số (c), r = số lần lặp lại, a = số nghiệm thức lô chính, b = số nghiệm thức lô phụ, c= số nghiệm thức lô phụ-trong lô phụ.

Đối với so sánh loại (8), (9),(10), (11) và (12), giá trị t' là giá trị t bảng có gia trọng được tính theo các công thức sau:

(Loại 8) $$t' = \frac{(b-1)E_b t_b + E_a t_a}{(b-1)E_b + E_a}$$

(Loại 9) $$t' = \frac{(c-1)E_c t_c + E_b t_b}{(c-1)E_c + E_b}$$

(Loại 10) $t' = \dfrac{(c-1)E_c t_c + E_b t_b}{(c-1)E_c + E_b}$

(Loại 11) $t' = \dfrac{(c-1)E_c t_c + E_a t_a}{(c-1)E_c + E_a}$

(Loại 12) $t' = \dfrac{b(c-1)E_c t_c + (b-1)E_b t_b + E_a t_a}{b(c-1)E_c + (b-1)E_b + E_a}$

Giá trị LSD_{DN} được tính như sau:

$$* \ \alpha = 0{,}05 \Rightarrow LSD_{DN} = 2{,}4525.\sqrt{\frac{(2).(0{,}095)}{4}} = 0{,}535$$

$$* \ \alpha = 0{,}01 \Rightarrow LSD_{DN} = 3{,}215.\sqrt{\frac{(2).(0{,}095)}{4}} = 0{,}701$$

So sánh giá trị sai khác giữa từng nghiệm thức và đối chứng với giá trị LSD_{DN} ; đánh dấu **, * hoặc ns trên mỗi trị số sai khác (Bảng 8.12)

Như vậy, với kiểm định Dunnet có hai nghiệm thức 5 và 6 khác biệt không ý nghĩa với đối chứng. Các nghiệm thức còn lại đều khác biệt ý nghĩa với đối chứng.

Bảng 8.12: So sánh năng suất trung bình của nghiệm thức đối chứng với một trong sáu nghiệm thức thuốc trừ sâu, dùng kiểm định LSD_{DN} (số liệu ở Bảng 5.1)

Nghiệm thức	Năng suất trung bình kg/ha (a)	Khác biệt với đối chứng kg/ha (b)
1. Dol-Mix (1kg)	2,127	0,811**
2. Dol-Mix (2kg)	2,678	1,362**
3. DDT + -BHC	2,552	1,236**
4. Azodrin	2,128	0,812**
5. Dimecron-Boom	1,796	$0{,}480^{ns}$
6. Dimecron-Knap	1,681	$0{,}365^{ns}$
7. Đối chứng	1,316	-

(a) Trung bình của 4 lần lặp lại

(b) ** = khác biệt ở mức ý nghĩa 1%, * = khác biệt ở mức ý nghĩa 5%, và ^{ns} = khác biệt không ý nghĩa

III. KIỂM ĐỊNH DUNCAN

1. Thí nghiệm một nhân tố

Đối với thí nghiệm cần đánh giá tất cả các cặp trung bình nghiệm thức, kiểm định LSD thường không thích hợp, đặc biệt khi số nghiệm thức lớn. Các bước tính trong kiểm định Duncan như sau:

Bước 1: Xếp trung bình nghiệm thức theo thứ tự tăng (hoặc giảm) dần.

Ví dụ: Bảy trung bình nghiệm thức (số liệu ở Bảng 5.1).được xếp theo thứ tự tăng dần như sau:

Nghiệm thức	Năng suất trung bình kg/ha	Xếp hạng
T7: Đối chứng	1,316	1
T6: Dimecron-Knap	1,681	2
T5: Dimecron-Boom	1,796	3
T1: Dol-Mix (1kg)	2,127	4
T4: Azodrin	2,128	5
T3: DDT + -BHC	2,552	6
T2: Dol-Mix (2kg)	2,678	7

Bước 2: Tính sai số chuẩn ($s_{\bar{X}}$) cho từng kiểu bố trí đặc biệt. Trong ví dụ, $s_{\bar{X}}$ được tính như sau:

$$s_{\bar{X}} = \sqrt{\frac{s_e^2}{r}} = \sqrt{\frac{0,095}{4}} = 0,154(t/ha)$$

Bước 3: Tính (t - 1) giá trị của các khoảng ngắn nhất có ý nghĩa như sau:

$R_p = r_p s_{\bar{X}}$ cho p = 2, 3, ..., t

với t là tổng số nghiệm thức, $s_{\bar{X}}$ là sai số chuẩn của sai khác trung bình được tính ở bước 2, các giá trị r_p là giá trị bảng đạt được từ Phụ lục X và p là khoảng cách theo thứ tự xếp hạng giữa các cặp trung bình nghiệm thức được so sánh (nghĩa là, p = 2 cho cặp trung bình có khoảng cách gần nhau nhất và p = t cho cặp trung bình có khoảng cách xa nhau nhất).

Trong ví dụ, các giá trị r_p trong Phụ lục X với độ tự do bằng 21 ở mức ý nghĩa 5% là:

p	$r_p(0,05)$
2	2,94
3	3,09
4	3,18
5	3,24
6	3,30
7	3,33

(t - 1) = 6 giá trị R_p được tính như sau:

p	$R_p = r_p s_{\bar{X}}$
2	(2,94)(0,154) = 0,453
3	(3,09)(0,154) = 0,476
4	(3,18)(0,154) = 0,490
5	(3,24)(0,154) = 0,499
6	(3,30)(0,154) = 0,509
7	(3,33)(0,154) = 0,513

Bước 4: Lập bảng trung bình đã được xếp thứ tự, với cột bên trái bỏ trị số lớn nhất và hàng trên cùng bỏ trị số nhỏ nhất. Trừ trung bình ở cột bên trái với trung bình của hàng trên cùng và ghi các sai khác vào trong bảng, tương tự như phương pháp kiểm định LSD khi so sánh tất cả các cặp trung bình nghiệm thức (xem Bảng 8.13).

Bảng 8.13: Các sai biệt của 7 trung bình mẫu đã được xếp hạng.

	1,681	1,796	2,127	2,128	2,552	2,678	So sánh với R_p	
1,316	0,365	0,480	0,811	0,812	1,236	1,362	⇐	0,513
1,681		0,115	0,446	0,447	0,871	0,997	⇐	0,509
1,796			0,331	0,332	0,756	0,882	⇐	0,499
2,127				0,001	0,425	0,551	⇐	0,490
2,128					0,424	0,550	⇐	0,476
2,552						0,126	⇐	0,453

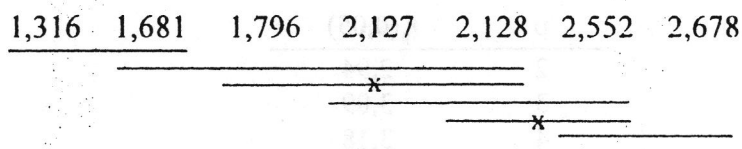
Chúng ta thấy đường chéo dài nhất (chứa các trị 0,365; 0,115;...; 0,126) là các sai biệt của hai trung bình có khoảng cách gần nhau nhất (p = 2). Đường chéo thứ hai gồm các sai biệt của hai trung bình cách nhau một trị trung bình (p = 3),..., và đường chéo ngắn nhất chỉ có một giá trị sai biệt của hai trung bình xa nhau nhất (nghĩa là, giữa giá trị nhỏ nhất và lớn nhất, p = t = 7)

So sánh các sai biệt trên từng đường chéo với trị số R_p tương ứng để tìm cặp trung bình khác biệt có ý nghĩa.

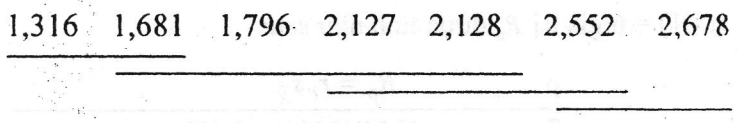
Bước 5: Trình bày kết quả kiểm định theo một trong hai cách:

(1) Dùng ký hiệu đoạn thẳng nếu trình bày kết quả theo thứ tự xếp hạng.

Trong dãy trung bình đã được xếp thứ tự, vẽ đoạn thẳng nối tất cả trung bình qua kiểm định Duncan khác biệt không ý nghĩa (tương tự như phương pháp kiểm định LSD khi so sánh tất cả các cặp trung bình nghiệm thức).

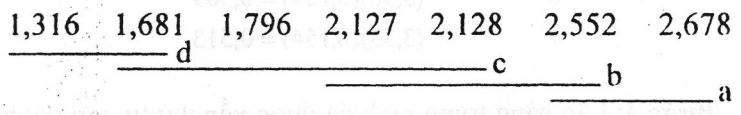


Vì đường thẳng thứ ba nằm trong đường thứ hai và đường thứ năm nằm trong đường thứ tư, nên có thể bỏ đường thứ ba và thứ năm. Kết quả sau cùng được trình bày như sau:



(2) Dùng ký hiệu alphabet nếu trình bày kết quả không theo thứ tự xếp hạng.

Ghi ký hiệu alphabet cho từng đoạn thẳng, thường chữ a được dùng đoạn thẳng chứa trung bình cao nhất, kế tiếp là b,... Ví dụ, ghi a, b, c, d cho bốn đoạn thẳng như sau:



Sau cùng, trình bày kết quả theo ký hiệu alphabet như ở Bảng 8.14. Chú ý, mỗi nghiệm thức có thể được ghi một hay nhiều chữ. Ví dụ, một chữ (a) cho nghiệm thức T2 và hai chữ (ab) cho nghiệm thức T3.

Bảng 8.14: Dùng kiểm định Duncan để so sánh tất cả các cặp trung bình nghiệm thức. Số liệu Bảng 5.1 của thí nghiệm CRD gồm 7 nghiệm thức (sử dụng ký hiệu alphabet)

Nghiệm thức	Năng suất trung bình kg/ha	Duncan
T1	2,127	bc
T2	2,678	a
T3	2,552	ab
T4	2,128	bc
T5	1,796	c
T6	1,681	cd
T7	1,316	d

Các trung bình có cùng chữ theo sau thì khác biệt không ý nghĩa ở mức $\alpha = 0.05$

2. Thí nghiệm hai hay nhiều nhân tố

Đối với thí nghiệm hai hay nhiều nhân tố, phương pháp kiểm DUNCAN được thực hiện tương tự như kiểm định LSD cho từng kiểu bố trí riêng biệt, nhưng thay giá trị t bảng bằng giá trị r_P của DUNCAN (Phụ lục X) và các giá trị $s_{\bar{d}}$ trong công thức tính LSD của các loại so sánh bằng giá trị $s_{\bar{x}}$, với

$$s_{\bar{x}} = \frac{s_{\bar{d}}}{\sqrt{2}}$$

Và các giá trị $R_P = r_P.s_{\bar{x}}$

IV. KIỂM ĐỊNH NEWMAN VÀ KEULS

Kiểm định Newman và Keuls cũng nhằm so sánh tất cả các cặp trung bình nghiệm thức. Các bước thực hiện cũng giống như kiểm định Duncan, ngoại trừ giá trị R_P được thay bằng giá trị R_{NK} của Newman và Keuls (Phụ lục XI)

$$R_{NK} = q_p.s_{\bar{x}}$$

với $s_{\bar{x}} = \sqrt{\frac{s_e^2}{r}}$ hoặc $s_{\bar{x}} = \frac{s_{\bar{d}}}{\sqrt{2}}$

s_e^2 = Trung bình bình phương sai số

r = Số lần lặp lại

q_P = Đại lượng được tìm thấy trong Bảng Newman và Keuls (Phụ lục XI)

V. PHƯƠNG PHÁP TƯƠNG PHẢN *(Contrast method)*

Phương pháp tương phản (so sánh trực giao = orthogonal comparison) là loại so sánh nhóm, thường được dùng để so sánh các trung bình nghiệm thức có dự định trước; nghĩa là, các contrast phải được lập từ lúc bắt đầu thí nghiệm. Phương pháp này cho phép chia nhỏ tổng bình phương nghiệm thức thành nhiều tổng bình phương có một độ tự do.

Nếu thí nghiệm có t nghiệm thức, có thể lập tối đa (t - 1) contrast trực giao lẫn nhau. Như vậy, tổng bình phương phương của tất cả các contrast sẽ bằng tổng bình phương nghiệm thức.

$SS(L_1) + SS(L_2) + \ldots + SS(L_{t-1})$ = SS(nghiệm thức)

1. Các nguyên tắc so sánh trực giao:

- **Nguyên tắc 1**: Một tổ hợp tuyến tính $L = \lambda_1 T_1 + \lambda_2 T_2 + \ldots + \lambda_t T_t$ được gọi là một contrast, với λ_i là hệ số contrast của nghiệm thức thứ i và T_i là tổng nghiệm thức thứ i. **Tổng các hệ số trong mỗi contrast phải bằng 0.**

$$\sum_{i=1}^{t} \lambda_i = 0$$

- **Nguyên tắc 2**: Hai contrast (L_1 và L_2) được xem là trực giao nếu tổng tích chéo của các hệ số của chúng bằng 0.

$$L_1 = \lambda_{12}\overline{X}_1 + \lambda_{12}\overline{X}_2 + \ldots + \lambda_{1t}\overline{X}_t$$
$$L_2 = \lambda_{21}\overline{X}_1 + \lambda_{22}\overline{X}_2 + \ldots + \lambda_{2t}\overline{X}_t$$

L_1 và L_2 được gọi là trực giao nếu

$$\sum \lambda_{1i}\lambda_{2i} = \lambda_{11}\lambda_{21} + \lambda_{12}\lambda_{22} + \cdots + \lambda_{1t}\lambda_{2t} = 0$$

2. Sai số chuẩn của một contrast

Sai số chuẩn của một contrast (L) được tính như sau:

$$\sqrt{\sum \lambda^2}(\sigma/\sqrt{n})$$

Sai số này được ước lượng bởi

$\sqrt{\sum \lambda^2}(s/\sqrt{n})$ với $s = \sqrt{s_e^2}$ (s_e^2 là trung bình bình phương sai số)

3. Tổng bình phương của một contrast

Tổng bình phương của một contrast L, SS(L)

$$SS(L) = \frac{L^2}{r(\sum \lambda^2)}$$

trong đó, r là số lần lặp lại (hay số khối)

Để kiểm định một contrast đặc biệt, chỉ cần kiểm định F của tổng bình phương của contrast đó với phương sai sai số (thật ra tổng bình phương của một contrast chính là phương sai, vì mỗi contrast chỉ có 1 độ tự do). Như vậy, chúng ta có một kiểm định F với độ tự do của tử số bằng 1 và độ tự do mẫu số chính là độ tự do của phương sai sai số. Phương pháp contrast cho phép trả lời các câu hỏi khá chính xác.

Ví dụ: Sử dụng số liệu của thí nghiệm về phân bón ở lúa với năm nghiệm thức (chương 5, Bảng 5.5).

Thí nghiệm có năm nghiệm thức; như vậy, sẽ có bốn so sánh độc lập có thể có. Giả sử, chúng ta có thể thực hiện các so sánh tương phản dưới đây:

- Ảnh hưởng của việc bón đạm (so sánh E với A-B-C-B)
- Ảnh hưởng của thời gian bón đạm lúc đẻ nhánh (A và B so với C và D)
- Ảnh hưởng của thời gian bón đạm lúc làm đòng (A và C so với B và D)
- Ảnh hưởng tương tác giữa thời gian bón đạm lúc đẻ nhánh và thời gian bón đạm lúc làm đòng (A và D so với B và C)

Các bước thực hiện:

Bước 1: Lập các contrast mong muốn và các hệ số (λ_i) của chúng. Chẳng hạn, chúng ta muốn thực hiện các contrast sau đây:

Ảnh hưởng	E	A	B	C	D	Tổng các hệ số
Nồng độ đạm	- 4	+ 1	+ 1	+ 1	+ 1	0
Giai đoạn đẻ nhánh	0	+ 1	+ 1	- 1	- 1	0
Giai đoạn làm đòng	0	+ 1	- 1	+ 1	- 1	0
Tương tác	0	+ 1	- 1	- 1	+ 1	0

Bước 2: Tính các tổng bình phương của 4 contrast

(1) Ảnh hưởng của việc bón đạm (so sánh E với A-B-C-B)

$$SS(L_1) = \frac{L^2}{r\left(\sum \lambda^2\right)} = \frac{(T_A + T_B + T_C + T_D - 4.T_{DC})^2}{4.[(+1)^2 + (+1)^2 + (+1)^2 + (+1)^2 + (-4)^2]}$$

$$= \frac{[(199,6 + 199,2 + 210,4 + 206,0) - (4)(152,0)]^2}{(4)(20)} = \frac{42931,84}{80} = 536,648$$

(2) Ảnh hưởng của thời gian bón đạm lúc đẻ nhánh (A và B so với C và D)

$$SS(L_2) = \frac{L^2}{r\left(\sum \lambda^2\right)} = \frac{[(T_A + T_B) - (T_C + T_D)]^2}{(+1)^2 + (+1)^2 + (-1)^2 + (-1)^2}$$

$$= \frac{[(199,6 + 199,2) - (210,4 + 206,0)]^2}{(4)(4)} = \frac{309,76}{16} = 19,36$$

(3) Ảnh hưởng của thời gian bón đạm lúc làm đòng (A và C so với B và D)

$$SS(L_3) = \frac{L^2}{r\left(\sum \lambda^2\right)} = \frac{[(T_A + T_C) - (T_B + T_D)]^2}{(+1)^2 + (+1)^2 + (-1)^2 + (-1)^2}$$

$$= \frac{[(199,6 + 210,4) - (199,2 + 206,0)]^2}{(4)(4)} = \frac{23,04}{16} = 1,44$$

(4) Ảnh hưởng tương tác giữa thời gian bón đạm lúc đẻ nhánh và thời gian bón đạm lúc làm đòng (A và D so với B và C)

$$SS(L_4) = \frac{L^2}{r\left(\sum \lambda^2\right)} = \frac{[(T_A + T_D) - (T_B + T_C)]^2}{(+1)^2 + (+1)^2 + (-1)^2 + (-1)^2}$$

$$= \frac{[(199,6 + 206,0) - (199,2 + 210,4)]^2}{(4)(4)} = \frac{166}{16} = 1$$

Bước 3: Lập bảng phân tích tổng bình phương nghiệm thức như sau:

Nguồn biến động	Độ tự do	Tổng bình phương	F (tính)	F (bảng) 5%	F (bảng) 1%
Nghiệm thức	4	558,448	49,15		
* Đạm	1	536,648	189,16**	4,75	9,33
* Thời gian bón lúc đẻ nhánh	1	19,36	6,82*		
* Thời gian bón lúc làn đòng	1	1,44	$0,52^{ns}$		
* Tương tác	1	1	$0,35^{ns}$		
Sai số	12	2,837			

Như vậy, có sự khác biệt ý nghĩa giữa đối chứng không bón đạm và các nghiệm thức có bón đạm; tương tự, giữa nghiệm thức bón đạm vào đầu và cuối đẻ nhánh. Trái lại, không có sự khác biệt ý nghĩa giữa nghiệm thức bón đạm vào đầu và cuối giai đoạn làm đòng cũng như tương tác giữa giai đoạn đẻ nhánh và giai đoạn làm đòng.

CHƯƠNG 9

PHÂN TÍCH HỒI QUI VÀ TƯƠNG QUAN

Trong phương pháp phân tích phương sai và so sánh các trung bình chỉ có thể đánh giá trên từng tính trạng riêng biệt; tuy nhiên, chúng ta biết rằng phản ứng của một tính trạng này có thể ảnh hưởng đến tính trạng khác hoặc tác động của một nghiệm thức có thể ảnh hưởng cùng lúc cả hai hay nhiều tính trạng. Chẳng hạn, chúng ta muốn nghiên cứu hai biến của cùng tổng thể: Tuổi và huyết áp, chiều cao và trọng lượng, chiều cao cây và năng suất,... Ngoài ra, cũng có thể có mối quan hệ nhiều biến; ví dụ, huyết áp ở người có liên quan đến tuổi, trọng lượng cơ thể và chế độ tập thể dục hoặc năng suất của một giống cây trồng có quan hệ với chiều cao cây, trọng lượng hạt, số trái trên cây,...

Phân tích tương quan và hồi qui cho phép khảo sát mối quan hệ giữa các biến. Phân tích tương quan cho số đo mức độ quan hệ giữa các biến, trong khi phân tích hồi qui mô tả ảnh hưởng của một hay nhiều biến *(biến độc lập)* trên một biến duy nhất *(biến phụ thuộc)* bằng cách biểu diễn biến phụ thuộc là một hàm số của biến độc lập. Để phân tích hồi qui, điều quan trọng cần phân biệt rõ giữa biến phụ thuộc và biến độc lập, mà sự phân biệt này không phải luôn luôn rõ ràng. Chẳng hạn phản ứng năng suất đối với đạm, năng suất dĩ nhiên là biến phụ thuộc và mức độ đạm là biến độc lập. Tuy nhiên, giữa năng suất hạt và hàm lượng protein thì sự nhận dạng các biến không rõ ràng. Thông thường tính trạng nào quan trọng hơn (như năng suất) sẽ trở thành biến phụ thuộc còn các yếu tố hoặc tính trạng ảnh hưởng đến năng suất sẽ trở thành biến độc lập.

Phương pháp phân tích tương quan và hồi qui có thể được xếp loại theo số biến khảo sát và dạng quan hệ hàm số giữa biến phụ thuộc và các biến độc lập. Phương pháp được gọi là đơn *(simple)* nếu chỉ có hai biến (một biến phụ thuộc và một biến độc lập), và bội *(multiple)* nếu có nhiều biến (một biến phụ thuộc và nhiều biến độc lập). Phương pháp được gọi là tuyến tính *(linear)* nếu dạng quan hệ cơ bản là đường thẳng; trái lại, được gọi là phi tuyến *(nonlinear)*. Do đó, phân tích tương quan và hồi qui có thể được xếp thành bốn loại:

1. Tương quan và hồi qui tuyến tính đơn
2. Tương quan và hồi qui tuyến tính bội
3. Tương quan và hồi qui phi tuyến đơn
4. Tương quan và hồi qui phi tuyến bội

A. QUAN HỆ TUYẾN TÍNH *(Linear Relationship)*

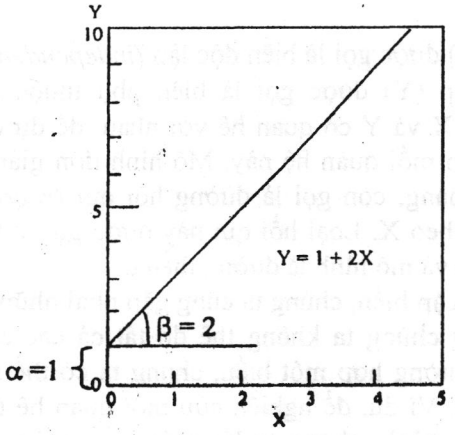
Hai biến có quan hệ tuyến tính nếu sự thay đổi là hằng số liên tiếp cả dãy số trong nghiên cứu. Đồ thị trình bày mối quan hệ tuyến tính là đường thẳng như được giải thích ở hình 9.1. Ở đây, Y gia tăng hằng số hai đơn vị cho mỗi đơn vị thay đổi ở X liên tiếp cả dãy giá trị X từ 0 đến 5: Y gia tăng từ 1 đến 3 khi X thay đổi từ 0 đến 1, và Y gia tăng từ 3 đến 5 khi X thay đổi từ 1 đến 2,...

Dạng hàm số của mối quan hệ tuyến tính giữa biến phụ thuộc Y và biến độc lập X được trình bày theo phương trình

$$Y = \alpha + \beta X$$

với α là tung độ gốc *(intercept)* của đường thẳng trên trục Y; và β, hệ số hồi qui tuyến tính, là độ dốc *(slope)* của đường thẳng hoặc số lượng thay đổi ở Y cho mỗi đơn vị thay đổi ở X. Ví dụ, đối với quan hệ tuyến tính của Hình 9.1, với tung độ gốc α bằng 1 và hệ số hồi qui tuyến tính β bằng 2, sự quan hệ được biểu diễn như sau:

$$Y = 1 + 2X \qquad \text{với } 0 \le X \le 5$$



Hình 9.1: Giải thích mối quan hệ tuyến tính giữa biến phụ thuộc Y và biến độc lập X

Với hai tham số của quan hệ tuyến tính (α và β) đã định, giá trị biến phụ thuộc Y, tương ứng với giá trị của biến độc lập X trong khoảng các giá trị X khảo sát có thể được xác định ngay lập tức bằng cách thay thế X trong phương trình với giá trị mong muốn và tính Y. Chẳng hạn, X = 1,5; Y được tính là 1 + 2(1,5) = 4; và với X = 3, Y được tính bằng 1 + 2(3) = 7.

Khi có nhiều hơn một biến độc lập, k biến độc lập ($X_1, X_2, \ldots, X_k$), dạng hàm số tuyến tính đơn của phương trình $Y = \alpha + \beta X$ có thể được triển khai sang dạng hàm tuyến tính bội là

$$Y = \alpha + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_k X_k$$

với α là tung độ gốc (nghĩa là, giá trị của Y khi tất cả giá trị X bằng 0) và β_i (i = 1, . . . k), hệ số hồi qui riêng phần (*partial regression coefficient*) liên quan đến biến độc lập X_i tiêu biểu cho số lượng thay đổi ở Y cho mỗi đơn vị thay đổi ở X_i. Do đó, ở dạng hàm tuyến tính bội với k biến độc lập sẽ (k + 1) tham số (nghĩa là α, $\beta_1, \beta_2, \cdots \beta_k$) cần được ước lượng.

Sự hiện diện của β_i (nghĩa là, khi giá trị β_i không bằng 0) chứng tỏ sự phụ thuộc của Y vào X_i. Mặt khác, nếu $\beta_i = 0$ lúc đó Y không phụ thuộc vào X_i (nghĩa là, không có sự quan hệ giữa Y và X_i theo dạng đã quy định). Vì vậy, kiểm định ý nghĩa của mỗi giá trị β_i để xác định β_i có bằng 0 hay không là phần cần thiết trong phân tích hồi qui.

Trong một vài trường hợp, nhà nghiên cứu cũng có thể mong muốn kiểm định ý nghĩa của tung độ gốc α (nghĩa là, kiểm tra $\alpha = \alpha_0$, với α_0 là giá trị được nhà nghiên cứu định rõ). Ví dụ, nếu nhà nghiên cứu mong muốn xác định Y = 0 khi giá trị của X trong phương trình $Y = \alpha + \beta X$ bằng 0, mà phương trình này tương đương với việc kiểm chứng đường thẳng có đi qua gốc tọa độ hay không, lúc đó ông sẽ kiểm định α = 0.

I. HỒI QUI TUYẾN TÍNH ĐƠN *(Simple Linear Regression)*

Câu hỏi thường đặt ra với cặp biến X và Y là: "Khi giá trị biến X thay đổi có làm ảnh hưởng đến giá trị của biến Y không?". Ví dụ, một em bé nam năm tuổi huyết áp sẽ như thế nào? Cũng có thể đặt một câu hỏi khác: " Giá trị mong đợi của Y đối với một giá trị nào đó của X là bao nhiêu?"; chẳng hạn, chúng ta muốn dự đoán huyết áp của nam ở tuổi 30.

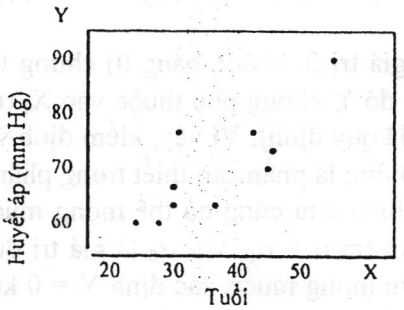
Biến tuổi (X) được gọi là biến độc lập *(independent)* hoặc biến dự đoán *(predictor)* và biến huyết áp (Y) được gọi là biến phụ thuộc *(dependent)* hoặc biến kết quả *(response)*. Nếu X và Y có quan hệ với nhau, để dự đoán Y từ X, chúng ta phải tìm một mô hình cho mối quan hệ này. Mô hình đơn giản nhất chỉ mối quan hệ giữa hai biến là đường thẳng, còn gọi là đường hồi qui *(regression line)*. Nghĩa là, chúng ta đang hồi qui Y theo X. Loại hồi qui này được gọi là hồi qui tuyến tính đơn vì chỉ có một biến độc lập và mô hình là đường thẳng.

Khi xét đến cặp biến, chúng ta cũng gặp phải những khó khăn giống như một biến; nghĩa là, thường chúng ta không thể đo tất cả các cá thể có thể có trong tổng thể. Tương tự như trường hợp một biến, chúng ta có thể sử dụng mẫu ngẫu nhiên để kết luận về tổng thể. Ví dụ, để nghiên cứu mối quan hệ tuyến tính giữa tuổi và huyết áp của nam trưởng thành, chúng ta lấy một mẫu ngẫu nhiên những người nam trưởng thành, đo huyết áp của họ để có được các số liệu mẫu về tuổi và huyết áp, sau đó xem đường thẳng có phù hợp với số liệu không?

Số liệu huyết áp đo được từ một mẫu ngẫu nhiên gồm 10 nam trưởng thành như sau:

Tuổi (X)	28	23	52	42	27	29	43	34	40	28
Huyết áp mmHg, (Y)	70	68	90	75	68	80	78	70	80	72

Chúng ta bắt đầu phân tích số liệu bằng cách chấm những điểm tương ứng với từng cặp (X, Y) trên đồ thị, với trục tung là giá trị của biến Y và trục hoành là giá trị của biến X (Hình 9.2). Những điểm này chắc chắn sẽ không nằm chính xác trên đường thẳng, nhưng dường như nó có dạng đường thẳng hướng lên; nghĩa là, tuổi càng cao thì huyết áp càng cao. Mô hình hồi qui được sử dụng thích hợp là đường thẳng đối với các số liệu như thế. Đây cũng là cách duy nhất để có thể dùng đường thẳng dự đoán huyết áp theo tuổi.



Hình 9.2: Đồ thị điểm phân tán của tuổi và huyết áp

Dĩ nhiên, hai biến cũng có thể có quan hệ khác hơn đường thẳng (quan hệ phi tuyến = nonlinear) hoặc có thể chúng không có quan hệ gì cả.

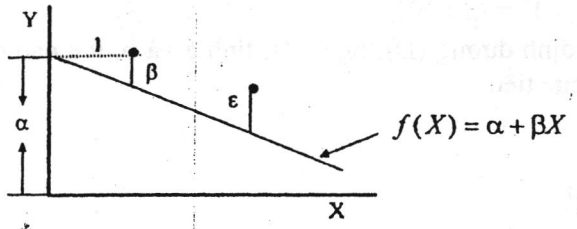
Mặc dầu hai biến có quan hệ đường thẳng, nhưng không phải tất cả các điểm quan sát sẽ nằm chính xác trên đường thẳng. Mô hình được viết dưới dạng toán học như sau:

$$Y = \alpha + \beta X + \varepsilon$$

Đường hồi qui được tính bởi hàm

$$f(X) = \alpha + \beta X$$

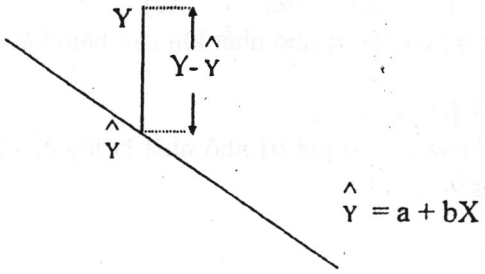
với α là tung độ gốc (intercept) và β là độ dốc (slope); nghĩa là, Y thay đổi theo mỗi đơn vị của X. ε chỉ độ lệch thẳng đứng từ một điểm quan sát đến đường hồi qui, đây là phần không được giải thích bởi mô hình nên được gọi là "sai số". Dạng đường thẳng của mô hình được trình bày qua Hình 9.3.



Hình 9.3: Đường hồi qui

1. Ước lượng α và β

Để tìm giá trị ước lượng chính xác nhất của α và β, chúng ta dùng phương pháp bình phương tối thiểu *(least squares)*, còn gọi là đường hồi qui mẫu *(sample regression line)*. Đường thẳng bình phương tối thiểu là đường thẳng duy nhất qua đó tổng bình phương của các khoảng lệch thẳng đứng từ điểm quan sát đến đường thẳng là nhỏ nhất (Hình 9.4).



Hình 9.4: Khoảng lệch thẳng đứng từ điểm quan sát đến đường bình phương tối thiểu.

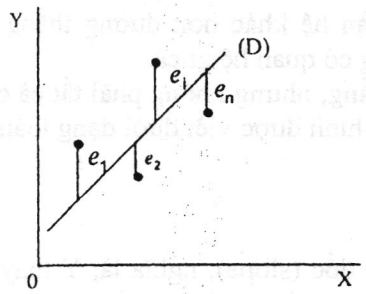
Giả sử, chúng ta có một tập hợp n điểm M_i có thể được đại diện bằng đường thẳng (D) có phương trình:

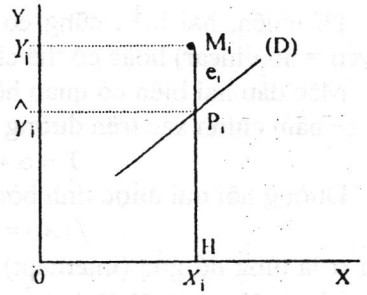
$$\hat{Y} = a + bX$$

với a là tung độ gốc của Y và b là độ dốc. Tối thiểu hóa hàm số:

$$f(a,b) = \sum_{i=1}^{n} (Y_i - \hat{Y})^2$$

trong đó, Y_i là giá trị quan sát và $\hat{Y}$ là giá trị được dự đoán từ đường thẳng đối với giá trị X tương ứng. Nghĩa là, chúng ta tìm a và b như thế nào để $(Y_i - \hat{Y})^2$ nhỏ nhất.





Mỗi điểm M_i có khoảng lệch e_i đối với đường (D), độ lệch e_i là đoạn M_iP_i.

$M_iP_i = M_iH - P_iH$

$e_i = Y_i - \hat{Y}_i$ với $\hat{Y}_i = (a + bX_i)$

Vấn đề đặt ra là xác định đường (D); nghĩa là, tính a và b; thế nào cho tổng bình phương các khoảng lệch cực tiểu.

$$A = \sum_{i=1}^{n} e_i^2$$
$$= \sum [Y_i - (a + bX_i)]^2$$
$$= \sum [Y_i^2 - 2Y_i(a + bX_i) + (a + bX_i)^2]$$
$$= \sum Y_i^2 - 2a\sum Y_i - 2b\sum X_iY_i + na^2 + 2ab\sum X_i + b^2\sum X_i^2$$

Chúng ta xem biểu thức trên như một tam thức bậc 2 đối với a:

$$na^2 - 2a(\sum Y_i - b\sum X_i) + \sum Y_i^2 - 2b\sum X_iY_i + b^2\sum X_i^2 \qquad (1)$$

hoặc đối với b:

$$b^2\sum X_i^2 - 2b(\sum X_iY_i - a\sum X_i) + \sum Y_i^2 - 2a\sum Y_i + na^2 \qquad (2)$$

Một tam thức bậc 2 ($\alpha X^2 + 2\beta X + \gamma$) có giá trị nhỏ nhất khi đạo hàm bậc nhất bằng 0 và đạo hàm bậc 2 dương, nghĩa là:

$$\alpha > 0 \text{ và } \alpha X + \beta = 0$$

Vậy điều kiện để hai tam thức (1) và (2) có giá trị nhỏ nhất là lấy đạo hàm riêng phần cho a và b, và cho đạo hàm bằng 0.

$$(1) \Rightarrow \frac{\partial A}{\partial a} = 2na - 2(\sum Y_i - b\sum X_i)$$
$$= na - \sum Y_i + b\sum X_i = 0$$
$$(2) \Rightarrow \frac{\partial B}{\partial b} = 2b\sum X_i^2 - 2(\sum X_iY_i - a\sum X_i)$$
$$= b\sum X_i^2 - \sum X_iY_i - a\sum X_i = 0$$

Chúng ta có hệ phương trình chuẩn (normal equation) như sau:

$$na + (\sum X_i)b = \sum Y_i \qquad (3)$$
$$(\sum X_i)a + (\sum X_i^2)b = \sum X_iY_i \qquad (4)$$

Với: $\overline{X} = \frac{\sum X_i}{n}$ và $\overline{Y} = \frac{\sum Y_i}{n}$

Chia (3) cho n:

$\Rightarrow a = \overline{Y} - b\overline{X}$

Thế a vào (4):

$$b = \frac{\sum X_i Y_i - [(\sum X_i \sum Y_i)/n]}{\sum X_i^2 - [(\sum X_i)^2/n]}$$

Mẫu số của độ dốc (b) chính là tổng bình phương của biến X (SS_{XX}):

Đặt: $\sum x^2 = \sum (X_i - \overline{X})^2 = \sum X_i^2 - \left[(\sum X_i)^2/n\right] = SS_{XX}$

Tử số của b là tổng tích chéo của hai biến X và Y (SS_{XY}):

Đặt: $\sum xy = \sum (X - \overline{X})(Y - \overline{Y}) = \sum XY - (\sum X)(\sum Y)/n = SS_{XY}$

Như vậy, độ dốc (b) có viết dưới dạng:

$$b = \frac{\sum (X_i - \overline{X})(Y_i - \overline{Y})}{\sum (X_i - \overline{X})^2}$$

hoặc $b = \dfrac{\sum xy}{\sum x^2}$ hoặc $b = \dfrac{SS_{XY}}{SS_{XX}}$

Phương pháp hồi qui có thể được sử dụng cho nhiều mục đích khác nhau:

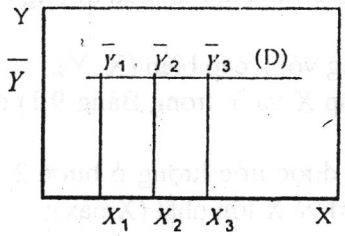
(1) Xem Y có phụ thuộc X không?

(2) Dự đoán Y theo X

(3) Nghiên cứu dạng của phương trình

(4) Dùng quan hệ giữa Y và X để kiểm soát biến động của Y

Để giải quyết mục đích (1), chúng ta đặt giả thuyết H_0 (không phụ thuộc) và nếu kết quả X và Y độc lập thì đường thẳng quan sát sẽ song song với trục X, nghĩa là, hệ số gốc (β) của đường thẳng bằng 0 (Hình 9.5). Trái lại, nếu hệ số gốc khác 0 nhiều, chúng ta bác bỏ giả thuyết H_0 và kết luận hai biến có quan hệ với nhau (Hình 9.3).



Hình 9.5: Hai biến X và Y độc lập

2. Tìm đường hồi qui ước lượng ($\hat{Y}$)

Ví dụ: Nghiên cứu mối quan hệ giữa đạm (N) và năng suất lúa. Chúng ta sử dụng số liệu từ thí nghiệm phân bón gồm bốn mức độ N. Số liệu được trình bày trong Bảng 9.1. Mục đích chính của phân tích hồi qui tuyến tính đơn là ước lượng sự đáp ứng tuyến tính của năng suất lúa với mức độ đạm được áp dụng, và để trắc nghiệm xem sự đáp ứng tuyến tính này có ý nghĩa hay không? Các bước thực hiện như sau:

Bước 1: Tính

$$\overline{X} = \frac{\sum X_i}{n} = \frac{300}{4} = 75 \quad \text{và} \quad \overline{Y} = \frac{\sum Y_i}{n} = \frac{23483}{4} = 5870{,}75$$

Bước 2: Tính a, b:

$$b = \frac{\sum xy}{\sum x^2} = \frac{249.475}{12.500} = 19{,}96$$

hoặc áp dụng công thức

$$b = \frac{\sum X_i Y_i - [(\sum X_i \sum Y_i)/n]}{\sum X_i^2 - [(\sum X_i)^2/n]} = \frac{2.010.700 - (300).(23.483)/4}{35.000 - (300)^2/4} = 19{,}96$$

$a = \overline{Y} - b\overline{X}$ = 5.870,75 - (19,96)(75) = 4.374

Do đó, đường hồi qui ước lượng là:

$\hat{Y}$ = a + bX = 4.374 + 19,96X với 0 < X < 150

Bảng 9.1: Tính phương trình hồi qui tuyến tính đơn giữa năng suất và mức độ đạm. Dùng số liệu của thí nghiệm phân bón trên lúa (Trích từ bảng 9.1 của Gomez and Gomez. 1984. Trang 363)

Mức độ đạm kg/ha (X)	Năng suất kg/ha (Y)	Lệch số từ trung bình x	Lệch số từ trung bình y	Lệch số bình phương x^2	Lệch số bình phương y^2	Tích các lệch số (x)(y)
0	4230	-75	-1640,75	5625	2692061	123056
50	5442	-25	-428,75	625	183827	10719
100	6661	25	790,25	625	624495	19756
150	7150	75	1279,25	5625	1636481	95944
			Tổng			
300	23483	0	0,00	12.500	5136864	249475
			Trung bình			
75	5870,75					

Bước 3: Chấm những điểm quan sát và vẽ đồ thị của đường hồi qui ước lượng tìm được ở bước 2 như sau:

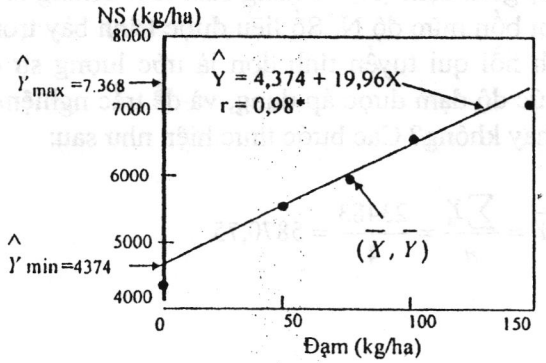
* Chấm n điểm quan sát, tương ứng với n cặp biến (X, Y).

Ví dụ: Bốn điểm quan sát (bốn cặp X và Y trong Bảng 9.1) được chấm trên Hình 9.6.

* Dùng đường hồi qui tuyến tính được ước lượng ở bước 2 để tính hai giá trị Y, tương ứng với trị số X nhỏ nhất (Xmin) và X lớn nhất (Xmax):

$\hat{Y}_{min}$ = a + b(X_{min}) = 4.374 + 19,96(0) = 4.374 kg/ha

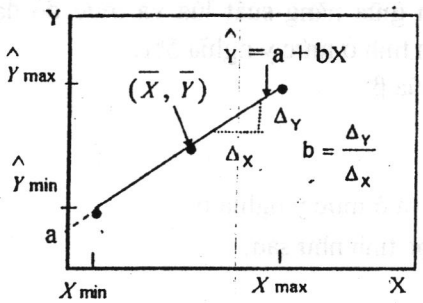
$\hat{Y}_{max}$ = a + b(X_{max}) = 4.374 + 19,96(150) = 7.368 kg/ha



Hình 9.6 : Đường hồi qui tuyến tính ước lượng giữa năng suất (Y) và mức độ đạm (X) được tính từ số liệu ở Bảng 9.1

* Chấm hai điểm ($X_{min}, \hat{Y}_{min}$) và ($X_{max}, \hat{Y}_{max}$) trên mặt phẳng (X, Y) và vẽ đoạn thẳng nối hai điểm này như Hình 9.7. Lưu ý các tính chất của đồ thị hồi qui tuyến ính:

(1) Đường thẳng chỉ có giá trị trong khoảng các trị số của X_{min} và X_{max}; ngoài khoảng này, đường thẳng sẽ không có giá trị.



Hình 9.7 : Đồ thị của đường hồi qui ước lượng : $\hat{Y} = a + bX$

(2) Đường thẳng phải đi qua điểm ($\bar{X}, \bar{Y}$).

(3) b là độ dốc (slope) của đường thẳng hay còn gọi là hệ số hồi qui.

(4) Đường thẳng nếu được kéo dài phải cắt trục Y tại trị số Y = a (tung độ gốc).

Bước 4: Kiểm định giả thuyết về β:

Sự hiện diện của β chứng tỏ có sự phụ thuộc của Y theo X, vì nếu β = 0, lúc đó Y không phụ thuộc vào X (nghĩa là, X và Y không có quan hệ tuyến tính). Do đó, trong phân tích hồi qui cần kiểm định ý nghĩa cho β để kết luận β = 0 hoặc β ≠ 0. Ngoài ra, chúng ta cũng có thể kiểm định β bằng một giá trị nào đó ($\beta = \beta_0$). Để kiểm định β cần tính:

* Trung bình bình phương sai số của biến Y theo biến X (s^2_{yx}):

$$s^2_{yx} = \frac{\sum(Y-\hat{Y})^2}{n-2}$$

hoặc $$s^2_{yx} = \frac{\sum(Y-\bar{Y})^2 - \left[\sum(X-\bar{X})(Y-\bar{Y})\right]^2 / \sum(X-\bar{X})^2}{n-2} = \frac{\sum y^2 - \left(\sum xy\right)^2 / \sum x^2}{n-2}$$

hoặc $$s^2_{yx} = \frac{\left[\sum Y^2 - \left(\sum Y\right)^2/n\right] - b.\left(\sum XY - \sum X \sum Y/n\right)}{n-2} = \frac{\sum y^2 - b\sum xy}{n-2}$$

Trong ví dụ, s^2_{yx} được tính như sau:

$$s^2_{yx} = \frac{5.136.864 - \frac{(249.475)^2}{12.500}}{4-2} = 78.921$$

* Tính trị số t_b:

$$t_b = \frac{b}{\sqrt{\frac{s^2_{yx}}{\sum X^2 - \left(\sum X\right)^2 / n}}}$$

$$t_b = \frac{19,96}{\sqrt{\frac{78.921}{12.500}}} = 7,94$$

* So sánh giá trị t_b tính với các trị số t bảng (Phụ lục IV) ở (n - 2) độ tự do. Kết luận $\beta \neq 0$ có ý nghĩa, nếu trị tuyệt đối của t_b tính lớn hơn trị số t bảng ở mức ý nghĩa α

Giá trị t bảng với df = (n - 2) = 2 ở mức ý nghĩa 5% là 4,303 và ở 1% là 9,925. Vì giá trị t_b tính lớn hơn giá trị t bảng ở mức ý nghĩa 5%, nhưng nhỏ hơn giá trị t bảng ở mức ý nghĩa 1%, nên chúng ta kết luận giữa năng suất lúa và mức độ đạm (trong khoảng từ 0 - 150 kg/ha) có quan hệ tuyến tính ở mức ý nghĩa 5%.

Bước 5: Tính khoảng tin cậy (1 - α) của β:

$$C.I.: b \pm t_\alpha \sqrt{\frac{s_{yx}^2}{\sum x^2}}$$

với t_α là trị số t bảng với (n - 2) độ tự do và ở mức ý nghĩa α.

Ví dụ: Khoảng tin cậy 95% cho β được tính như sau:

$$\text{C.I.}_{0,95}: \; b \pm t_{0,05}\sqrt{\frac{s_{yx}^2}{\sum X^2 - \left(\sum X\right)^2 / n}}$$

$$19,96 \pm 4,303\sqrt{\frac{78.921}{12.500}}$$

$$19,96 \pm 10,81$$

$$9,15 < \beta < 30,77$$

Như vậy, trong khoảng 0 - 150 kg N/ha, chúng ta có thể dự đoán mỗi lần gia tăng 1 kg N/ha, năng suất hạt sẽ tăng từ 9,15 đến 30,77 kg/ha, với độ tin cậy là 95%.

Bước 6: Kiểm định giả thuyết $\alpha = \alpha_0$:

Thường không cần thiết kiểm định α; tuy nhiên, trong một vài trường hợp, thí nghiệm viên cũng muốn kiểm định ý nghĩa của α; nghĩa là, muốn kiểm chứng $\alpha = \alpha_0$ hoặc $\alpha \neq \alpha_0$ với α_0 là trị số nào đó do thí nghiệm viên chọn. Cách tính như sau:

* Tính trị số t_a:

$$t_a = \frac{a - \alpha_0}{s_{yx}^2[(1/n) + (\overline{X}^2 / \sum x^2)]}$$

* So sánh giá trị t_a với giá trị t bảng (Phụ lục IV), với (n - 2) độ tự do ở mức ý nghĩa α. Loại giả thuyết $\alpha = \alpha_0$ nếu giá trị tuyệt đối của t_a lớn hơn giá trị t bảng tương ứng.

Ví dụ: Chúng ta muốn kiểm định α (nghĩa là, năng suất ở 0 kg N/ha) có khác biệt ý nghĩa với 4000 kg/ha không? Trị số t_a được tính như sau:

$$t_a = \frac{4.374 - 4.000}{78.921\left[\frac{1}{4} + \frac{(75)^2}{12.500}\right]} = 1,59$$

Giá trị t bảng, với (n - 2) độ tự do, ở mức ý nghĩa 5% là 4,303. Vì giá trị t_a nhỏ hơn giá trị t bảng, nên chúng ta có thể kết luận giá trị α khác biệt không ý nghĩa với 4000 kg/ha.

II TƯƠNG QUAN TUYẾN TÍNH ĐƠN *(Simple Linear Correlation)*

Mục đích của phân tích tương quan tuyến tính đơn là ước lượng và kiểm định ý nghĩa hệ số tương quan tuyến tính đơn r. Đây là số đo mức độ quan hệ giữa hai biến X

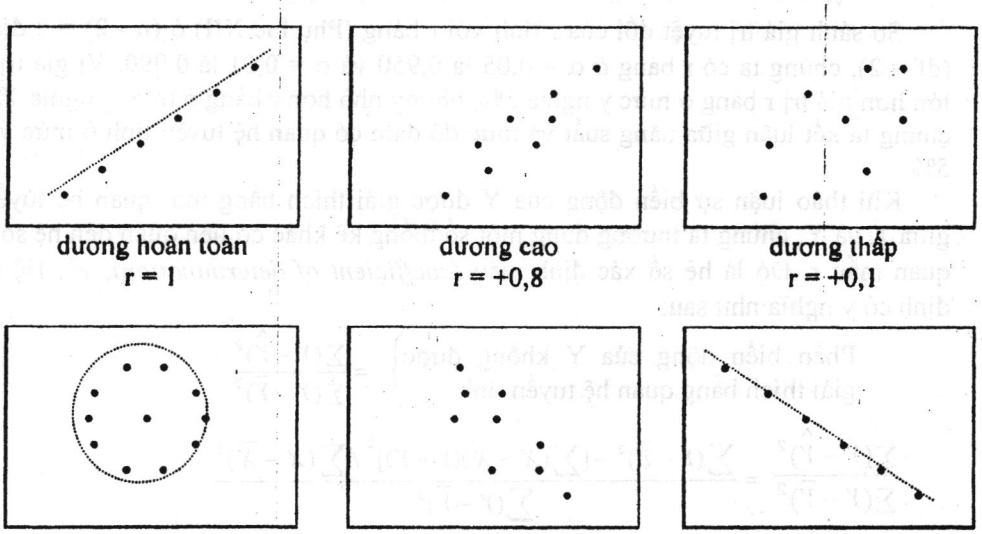
và Y. Hệ số tương quan tuyến tính đơn được tính dựa trên số lượng biến động của một biến mà có thể được giải thích bằng hàm tuyến tính của biến khác. Kết quả sẽ giống nhau nếu Y được biểu diễn như hàm tuyến tính của X hoặc X được biểu diễn như hàm tuyến tính của Y. Do đó, khi tính hệ số r không cần để ý biến nào là nguyên nhân (biến độc lập), biến nào là hậu quả (biến phụ thuộc) như yêu cầu trong phân tích hồi qui.

Công thức tính hệ số tương quan như sau:

$$r = \frac{\sum (X-\overline{X})(Y-\overline{Y})}{\sqrt{\sum (X-\overline{X})^2 \sum (Y-\overline{Y})^2}} = \frac{\sum xy}{\sqrt{\sum x^2 . \sum y^2}}$$

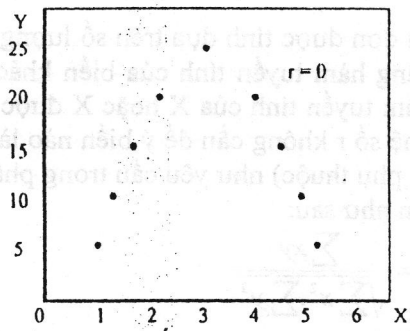
hoặc $$r = \frac{\sum XY - [(\sum X)(\sum Y)/n]}{\sqrt{\left[\sum X^2 - \frac{(\sum X)^2}{n}\right]\left[\sum Y^2 - \frac{(\sum Y)^2}{n}\right]}}$$

Không giống như hệ số hồi qui (β), hệ số tương quan không có đơn vị đo lường. Do đó, dựa vào độ lớn của trị tuyệt đối r, chúng ta có thể biết được độ mạnh của quan hệ tuyến tính giữa hai biến. Trong tất cả các trường hợp r biến động giữa - 1 và + 1 ($-1 \le r \le +1$). Nếu r = - 1, X và Y có sự tương quan nghịch chiều hoàn toàn và tất cả các số liệu quan sát sẽ nằm trên đường hồi qui mẫu với độ dốc âm. Nếu r = + 1, tương quan cùng chiều hoàn toàn và tất cả số liệu quan sát nằm trên đường hồi qui mẫu với độ dốc dương. Khi r gần bằng 0, hầu như không có sự quan hệ giữa hai biến. Vì vậy, nhìn dấu và độ lớn của r có thể đoán được chiều, độ mạnh của mối quan hệ (Hình 9.8).



Hình 9.8: Đồ thị của các hệ số tương quan r có giá trị khác nhau.

Mặc dầu r = 0 chứng tỏ không có sự tương quan tuyến tính giữa hai biến, nhưng nó không có nghĩa là không có sự tương quan nào giữa chúng. Có thể hai biến có tương quan phi tuyến *(nonlinear relationship)*, như tương quan bậc 2 có r = 0 (Hình 9.9). Đây là lý do tại sao chúng ta dùng từ hệ số tương quan tuyến tính đơn, thay vì dùng tên gọi thông thường là hệ số tương quan đơn hoặc hệ số tương quan.



Hình 9.9: Tương quan bậc 2 giữa hai biến X và Y. Kết quả tính hệ số tương quan tuyến tính đơn bằng 0.

Ví du: Sử dụng số liệu trong phân tích hồi qui tuyến tính đơn. Chúng ta áp dụng phân tích tương quan tuyến tính đơn để xác định độ mạnh của quan hệ tuyến tính giữa năng suất và mức độ đạm, giá trị r được tính như sau:

$$r = \frac{\sum XY - [(\sum X)(\sum Y)/n]}{\sqrt{\left[\sum X^2 - \frac{(\sum X)^2}{n}\right]\left[\sum Y^2 - \frac{(\sum Y)^2}{n}\right]}} \quad \text{hoặc} \quad r = \frac{\sum xy}{\sqrt{\sum x^2 \sum y^2}}$$

$$r = \frac{249.475}{\sqrt{(12.500)(5.136.864)}} = 0{,}985$$

hoặc

$$r = \frac{2.010.700 - (300).(23.483)/4}{\sqrt{[35.000 - (300)^2/4][142.999.685 - (23.483)^2/4]}} = 0{,}985$$

So sánh giá trị tuyệt đối của r tính với r bảng (Phụ lục XII) ở (n - 2) = 2 độ tự do (df = 2), chúng ta có r bảng ở $\alpha = 0{,}05$ là 0,950 và $\alpha = 0{,}01$ là 0,990. Vì giá trị r tính lớn hơn giá trị r bảng ở mức ý nghĩa 5%, nhưng nhỏ hơn r bảng ở mức ý nghĩa 1%, nên chúng ta kết luận giữa năng suất và mức độ đạm có quan hệ tuyến tính ở mức ý nghĩa 5%.

Khi thảo luận sự biến động của Y được giải thích bằng mối quan hệ tuyến tính giữa X và Y, chúng ta thường dùng một số thống kê khác có liên quan đến hệ số tương quan mẫu r. Đó là hệ số xác định mẫu *(coefficient of determination)*, r^2. Hệ số xác định có ý nghĩa như sau:

$$\left.\begin{array}{l}\text{Phần biến động của Y không được}\\ \text{giải thích bằng quan hệ tuyến tính}\end{array}\right\} = \frac{\sum(Y-\hat{Y})^2}{\sum(Y-\bar{Y})^2}$$

$$\frac{\sum(Y-\hat{Y})^2}{\sum(Y-\bar{Y})^2} = \frac{\sum(Y-\bar{Y})^2 - [\sum(X-\bar{X})(Y-\bar{Y})]^2/\sum(X-\bar{X})^2}{\sum(Y-\bar{Y})^2}$$

$$= 1 - \frac{(\sum xy)^2}{(\sum x^2)(\sum y^2)} = 1 - r^2$$

vì thế, $r^2 = 1$ - phần biến động không được giải thích trong tổng thể.

Như vậy, r^2 chỉ phần biến động của Y được giải thích bằng quan hệ tuyến tính với X. Nếu r^2 lớn (≈ 1) hầu hết biến động được giải thích bằng mối quan hệ tuyến tính. Nếu $r^2 \approx 0$, hầu như không có quan hệ tuyến tính giữa hai biến và thông tin về độ lớn của biến X cho rất ít thông tin về độ lớn của biến Y. r^2 là số thống kê có ý nghĩa nhất, ngay cả trong phân tích hồi qui, nó thường là số thống kê được người làm thí nghiệm

quan tâm đầu tiên để xác định xem phương trình hồi qui có giúp ích cho việc dự đoán biến Y?

Trong ví dụ, chúng ta tính được r = 0,985; chứng tỏ rằng $(100)(0,985)^2 = 97\%$ ($= r^2$) của biến động ở năng suất trung bình được giải thích bởi hàm tuyến tính của mức độ đạm được áp dụng. Giá trị r đạt được tương đối cao cũng chứng tỏ tính chặt chẽ giữa đường hồi qui ước lượng và các điểm quan sát (Hình 9.5). Tương quan tuyến tính giữa năng suất trung bình và mức độ đạm được áp dụng dường như phù hợp với số liệu một cách thỏa đáng trong khoảng từ 0 đến 150 kg N/ha.

Chú ý, n càng nhỏ thì giá trị r (tính) phải càng lớn, kết luận mới có ý nghĩa. Do đó, trong thực tế độ ý nghĩa và độ lớn của r phải được công bố cùng với cỡ mẫu (n).

Vì trong bảng hệ số tương quan tuyến tính đơn (Phụ lục XII) chỉ cho các giá trị r ở hai mức ý nghĩa: 5% và 1% và độ tự do đến 500 Trong trường hợp độ tự do (n - 2) lớn hơn 500 và xác suất α nhỏ hơn 0,01, chúng ta áp dụng công thức sau đây để kiểm định hệ số r:

$$t = \frac{r}{\sqrt{1 - r^2}} \sqrt{n - 2}$$

Dùng bảng t với xác suất α tương ứng ở độ tự do (n - 2).

III. TÍNH ĐỒNG NHẤT CỦA CÁC HỆ SỐ HỒI QUI

Trong thí nghiệm một nhân tố, ở đó chỉ có một nhân tố được phép biến đổi, trong khi đối với thí nghiệm thừa số, có nhiều nhân tố được phép biến đổi. Ví dụ, với số liệu của thí nghiệm hai nhân tố bao gồm bốn giống lúa và năm mức độ phân đạm, hồi qui tuyến tính của năng suất trên mức phân đạm có lẽ phải được khảo sát riêng biệt cho mỗi giống. Hoặc với số liệu của thí nghiệm ba nhân tố bao gồm ba giống, bốn mật độ gieo và năm mức phân đạm, 12 hồi qui riêng biệt giữa năng suất và các mức độ đạm cần được ước lượng cho một trong 3 x 4 = 12 phối hợp nghiệm thức của giống và mật độ gieo. Tương tự, nếu chúng ta quan tâm đến việc khảo sát mối quan hệ giữa năng suất và mật độ gieo, có thể ước lượng phương trình hồi qui cho một trong 3 x 5 = 15 phối hợp nghiệm thức của giống và đạm.

Theo cách tương tự, đối với các thí nghiệm trong một loạt môi trường (nghĩa là, ở các địa điểm hoặc mùa vụ hoặc năm khác nhau), phân tích hồi qui cần được áp dụng riêng biệt cho mỗi thí nghiệm.

Khi ước lượng nhiều hồi qui tuyến tính, thường chúng ta chú ý để xác định xem các hệ số hồi qui (hay độ dốc) của các đường hồi qui có khác nhau hay không. Ví dụ, trong thí nghiệm hai nhân tố bao gồm giống và mức phân đạm, điều quan trọng cần biết là mức độ thay đổi năng suất khi tăng mức phân đạm có thay đổi giữa giống này và giống khác hay không? Câu hỏi như thế sẽ được trả lời bằng cách so sánh các hệ số hồi qui của các giống khác nhau. Điều này liên quan đến *kiểm định tính đồng nhất của các hệ số hồi qui.*

Khái niệm tính đồng nhất của các hệ số hồi qui có liên quan gần với khái niệm tương tác giữa các nhân tố đã được thảo luận ở chương thí nghiệm nhiều nhân tố. Các đường hồi qui có độ dốc bằng nhau thì song song với nhau, mà điều này cũng có nghĩa là không có sự tương tác giữa các nhân tố khảo sát. Nói cách khác, phản ứng đối với các mức độ của nhân tố A vẫn giống nhau trên các mức độ của nhân tố B.

Cũng cần lưu ý, tính đồng nhất của các hệ số hồi qui không hàm ý các đường hồi qui tương đương nhau. Đối với hai hay nhiều đường hồi qui để trùng khớp nhau (các

đường nằm chồng lên nhau), các hệ số hồi qui β và hệ số gốc α phải đồng nhất. Trong nghiên cứu nông nghiệp, phân tích hồi qui thường được áp dụng đối với các số liệu từ các thí nghiệm có kiểm soát; nhìn chung, các nhà nghiên cứu quan tâm nhiều đến việc so sánh các mức độ thay đổi (β) hơn là tung độ gốc α. Tuy nhiên, nếu nhà nghiên cứu mong muốn xác định xem có thể sử dụng một đường hồi qui duy nhất để đại diện cho nhiều đường hồi qui có các hệ số hồi qui đồng nhất, cách so sánh thích hợp các trung bình nghiệm thức (ở mức X = 0) có thể được thực hiện theo các phương pháp so sánh nhiều trung bình ở chương 8. Nếu sự khác biệt giữa các trung bình này không ý nghĩa, lúc đó có thể sử dụng một đường hồi qui duy nhất.

Phương pháp kiểm định tính đồng nhất của các hệ số hồi qui được trình bày dưới đây cho hai trường hợp, một trường hợp chỉ có hai hệ số hồi qui và một trường hợp khác có nhiều hơn hai hệ số hồi qui.

1. Hai hệ số hồi qui

Số liệu năng suất hạt (Y) và số chồi (X) ở Bảng 9.2 được dùng để giải thích phương pháp kiểm định giả thuyết $\beta_1 = \beta_2$ của hai đường hồi qui được tiêu biểu bởi hai phương trình: $Y_1 = \alpha_1 + \beta_1 X$ và $Y_2 = \alpha_2 + \beta_2 X$. Mục đích là để xác định xem các hệ số hồi qui trong quan hệ tuyến tính giữa năng suất và số chồi có giống nhau cho cả hai giống không? Các bước được thực hiện như sau:

Bảng 9.2 Tính hai hệ số hồi qui tuyến tính giữa năng suất hạt (Y) và số chồi (X) cho giống lúa Milfor 6(2) và Taichung Native 1 (Trích từ Bảng 9.3 của Gomez and Gomez. 1984. Trang 373)

Milfor 6(2)		Taichung Native 1	
Năng suất hạt (kg/ha)	Số chồi/m²	Năng suất hạt (kg/ha)	Số chồi/m²
4862	160	5380	293
5244	175	5510	325
5128	192	6000	332
5052	195	5840	342
8298	238	6416	342
5410	240	6666	378
5234	252	7016	380
5608	282	6994	410
$\overline{X}_1 = 217$		$\overline{X}_2 = 350$	
$\overline{Y}_1 = 5230$		$\overline{Y}_2 = 6228$	
$\sum x_1^2 = 12542$		$\sum x_2^2 = 9610$	
$\sum y_1^2 = 357630$		$\sum y_2^2 = 2872044$	
$\sum x_1 y_1 = 57131$		$\sum x_2 y_2 = 153854$	
$b_1 = \frac{57131}{12542} = 4{,}56$		$b_2 = \frac{153854}{9610} = 16{,}01$	
$a_1 = 5230 - (4{,}56)(217) = 4{,}240$		$a_2 = 6228 - (16{,}01)(350) = 624$	

Ghi chú: $\sum x_i^2 = \sum X_i^2 - \left(\sum X_i\right)^2 / n_i$

$$\sum y_i^2 = \sum Y_i^2 - \left(\sum Y_i\right)^2 / n_i$$

$$\sum x_i y_i = \sum X_i Y_i - \left(\sum X_i \sum Y_i\right) / n_i$$

Bước 1: Ước lượng hai đường hồi qui tuyến tính đơn cho từng giống:

$$\hat{Y}_1 = a_1 + b_1X$$

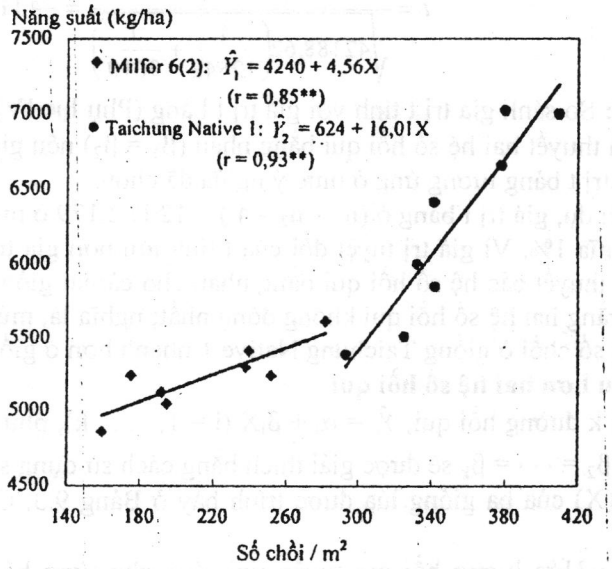
$$\hat{Y}_2 = a_2 + b_2X$$

Với số liệu ở Bảng 9.2, hai đường hồi qui ước lượng tính được là:

$\hat{Y}_1 = 4240 + 4{,}56X$ cho giống Milfor 6(2)

$\hat{Y}_2 = 624 + 16{,}01X$ cho giống Taichung Native 1

Hai đường hồi qui ước lượng này được trình bày bằng đồ thị ở **Hình 9.10.**



Hình 9.10: Các đường hồi qui tuyến tính ước lượng giữa năng suất (Y) và số chồi (X) của hai giống lúa: Milfor 6(2) và Taichung Native 1.

Bước 2: Tính trung bình bình phương sai số *(residual MS)* cho mỗi bộ số liệu:

$$s_{y.x}^2(1) = \frac{357630 - \frac{(57131)^2}{12542}}{8-2} = 16231{,}39$$

$$s_{y.x}^2(2) = \frac{2872044 - \frac{(153854)^2}{9610}}{8-2} = 68145{,}85$$

Bước 3: Tính trung bình bình phương sai số gộp (s_p) như sau:

$$s_p^2 = \frac{(n_1 - 2)s_{y.x}^2(1) + (n_2 - 1)s_{y.x}^2(2)}{n_1 + n_2 - 4}$$

với $s^2_{y.x}(1)$ và $s^2_{y.x}(2)$ là các giá trị của trung bình bình phương sai số được tính ở bước 2, n_1 và n_2 là số cặp quan sát của bộ số liệu thứ nhất và thứ hai. Nếu $n_1 = n_2$ thì s^2_p đơn giản là trung bình cộng của hai trung bình bình phương sai số. Trong ví dụ, vì $n_1 = n_2 = 8$ nên trung bình bình phương sai số được tính như sau:

$$s^2_p = \frac{16231,39 + 68145,85}{2} = 42188,62$$

Bước 4: Tính giá trị t theo công thức sau:

$$t = \frac{b_1 - b_2}{\sqrt{s^2_p\left(\frac{1}{\sum x_1^2} + \frac{1}{\sum x_2^2}\right)}}$$

với b_1 và b_2 là các hệ số hồi qui ước lượng.

Trong ví dụ, giá trị t được tính như sau:

$$t = \frac{4,56 - 16,01}{\sqrt{42188,62\left(\frac{1}{12542} + \frac{1}{9610}\right)}} = -4,11$$

Bước 5: So sánh giá trị t tính với giá trị t bảng (Phụ lục IV) với $(n_1 + n_2 - 4)$ độ tự do. Loại giả thuyết hai hệ số hồi qui bằng nhau ($\beta_1 = \beta_2$) nếu giá trị tuyệt đối của t tính lớn hơn giá trị t bảng tương ứng ở mức ý nghĩa đã chọn.

Trong ví dụ, giá trị t bảng ở $(n_1 + n_2 - 4) = 12$ là 2,179 ở mức ý nghĩa 5% và 3,055 ở mức ý nghĩa 1%. Vì giá trị tuyệt đối của t tính lớn hơn giá trị t bảng ở mức ý nghĩa 1% nên giả thuyết các hệ số hồi qui bằng nhau cho cả hai giống bị loại. Do đó, chúng ta kết luận rằng hai hệ số hồi qui không đồng nhất; nghĩa là, mức độ gia tăng năng suất hạt khi tăng số chồi ở giống Taichung Native 1 nhanh hơn ở giống Milfor 6(2).

2. Nhiều hơn hai hệ số hồi qui

Đối với k đường hồi qui, $\hat{Y}_i = \alpha_i + \beta_i X$ (i = 1, . . . , k), phương pháp kiểm định giả thuyết $\beta_1 = \beta_2 = \cdots = \beta_k$ sẽ được giải thích bằng cách sử dụng số liệu năng suất hạt (Y) và số chồi (X) của ba giống lúa được trình bày ở Bảng 9.3. Các bước thực hiện như sau:

Bước 1: Ước lượng hồi qui tuyến tính đơn cho từng bộ số liệu để đạt được k phương trình hồi qui tuyến tính:

$$\hat{Y}_i = a_i + b_i X \quad \text{với } i = 1, \ldots, k$$

Trong ví dụ, các phương trình hồi qui tuyến tính được ước lượng giữa năng suất hạt (Y) và số chồi (X) cho k = 3 giống như sau:

$\hat{Y}_1 = 1.436 + 536,70X$ cho giống IR1514A-E666

$\hat{Y}_2 = -3.414 + 860,70X$ cho giống IR8

$\hat{Y}_3 = 8.771 - 942,38X$ cho giống peta

Ba phương trình hồi qui ước lượng này được trình bày bằng đồ thị ở Hình 9.11.

Bước 2: Tính giá trị tổng bình phương sai số cho từng bộ số liệu:

$$A_i = \sum y_i^2 - \frac{\left(\sum xy\right)^2}{\sum x^2} \quad \text{với } i = 1, \ldots, k$$

Trong ví dụ, ba giá trị của tổng bình phương sai số cho các giống IR1514A-E666, IR8 và Peta là:

$$A_1 = 4068187 - \frac{(4519)^2}{8,42} = 1642847$$

$$A_2 = 8576031 - \frac{(9,373)^2}{10,89} = 508710$$

$$A_3 = 6493098 - \frac{(-4514)^2}{4,79} = 2239195$$

Bảng 9.3: Phép tính để kiểm định tính đồng nhất của ba hệ số hồi qui tuyến tính giữa năng suất hạt (Y) và số chồi (X) của ba giống IR1514A-E666, IR8 và Peta (Trích từ bảng 9.4 của Gomez and Gomez. 1984. Trang 377).

IR1514A-E666		IR8		Peta	
Năng suất kg/ha	Số chồi/hốc	Năng suất kg/ha	Số chồi/hốc	Năng suất kg/ha	Số chồi/hốc
5932	7,98	4876	9,98	1528	7,90
4050	5,72	3267	7,67	2858	6,39
4164	4,95	2051	6,67	3857	5,78
4862	7,82	4322	8,33	3796	5,55
5596	6,67	4557	9,16	1507	6,50
5570	6,55	4832	9,67	2078	7,67
4002	5,28	2322	6,75	1638	7,03
$\overline{X}_1 = 6,42$		$\overline{X}_2 = 8,32$		$\overline{X}_3 = 6,69$	
$\overline{Y}_1 = 4882$		$\overline{Y}_2 = 3747$		$\overline{Y}_3 = 2466$	
$\sum x_1^2 = 8,42$		$\sum x_2^2 = 10,89$		$\sum x_3^2 = 4,79$	
$\sum y_1^2 = 4068187$		$\sum y_2^2 = 8576031$		$\sum y_3^2 = 6493098$	
$\sum x_1 y_1 = 4519$		$\sum x_2 y_2 = 9373$		$\sum x_3 y_3 = -4514$	

Ghi chú: $\sum x_i^2 = \sum X_i^2 - (\sum X_i)^2 / n_i$

$$\sum y_i^2 = \sum Y_i^2 - (\sum Y_i)^2 / n_i$$

$$\sum x_i y_i = \sum X_i Y_i - (\sum X_i \sum Y_i) / n_i$$

Bước 3: Tính tổng của k = 3 giá trị của tổng bình phương sai số được tính ở bước 2 như sau:

$$B = \sum_{i=1}^{k} A_i$$

B = 1642847 + 508710 + 2239195 = 4390752

Bước 4: Tính tổng của k = 3 tổng bình phương đã hiệu chỉnh cho một trong hai biến (X và Y) và tổng của k tổng tích chéo đã hiệu chỉnh giữa hai biến theo công thức sau:

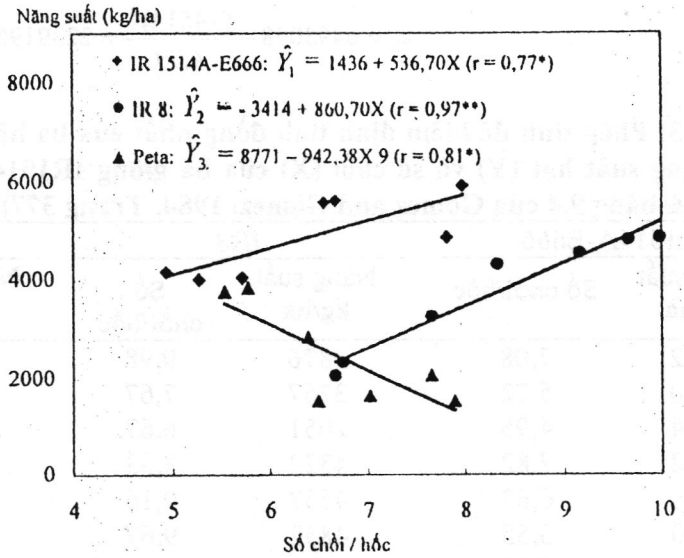
$$C = \sum x_1^2 + \sum x_2^2 + \cdots + \sum x_k^2$$

$$D = \sum y_1^2 + \sum y_2^2 + \cdots + \sum y_k^2$$

$$E = \sum x_1 y_1 + \sum x_2 y_2 + \cdots + \sum x_k y_k$$

Trong ví dụ, các giá trị C, D và E là:

$$C = 8,42 + 10,89 + 4,79 = 24,10$$
$$D = 4068187 + 8576031 + 6493098 = 19137316$$
$$E = 4519 + 9373 + (-4514) = 9378$$



Hình 9.11: Các đường hồi qui tuyến tính ước lượng giữa năng suất (Y) và số chồi (X) của ba giống lúa: IR 1514A-E666, IR 8 và Peta.

Bước 5: Tính giá trị F

$$F = \frac{\left[D - (E^2/C) - B\right]/(k-1)}{B\Big/\left(\sum_{i=1}^{k} n_i - 2k\right)}$$

với n_i (i = 1,..., k) là số cặp quan sát trong bộ số liệu thứ i. Trong ví dụ, giá trị F là:

$$F = \frac{\left\{19137316 - \left[(9378)^2/(24,10)\right] - 4390752\right\}/(3-1)}{4390752/\left[(7+7+7) - 2(3)\right]} = 18,96$$

Bước 6: So sánh giá trị F tính với giá trị F bảng với các độ tự do $df_1 = (k - 1)$ và $df_2 = (\sum_{i=1}^{k} n_i - 2k)$. Loại bỏ tính đồng nhất của k hệ số hồi qui ($\beta_1 = \beta_2 = \cdots = \beta_k$) nếu giá trị F tính lớn hơn giá trị F bảng tương ứng ở mức ý nghĩa đã chọn.

Trong ví dụ, giá trị F bảng với $df_1 = 2$ và $df_2 = 15$ là 3,68 ở mức ý nghĩa 5% và 6,36 ở mức ý nghĩa 1%. Vì giá trị F tính lớn hơn giá trị F bảng ở mức ý nghĩa 1% nên tính đồng nhất giữa ba hệ số hồi qui bị loại.

IV. TÍNH ĐỒNG NHẤT CỦA HỆ SỐ TƯƠNG QUAN

Để xác định xem mức độ quan hệ tuyến tính giữa hai biến vẫn giống nhau ở các mức độ khác nhau của biến thứ ba hoặc thứ tư, có thể thực hiện kiểm định tính đồng

nhất của các hệ số tương quan tuyến tính đơn. Ví dụ, các hệ số tương quan tuyến tính đơn giữa năng suất hạt và trọng lượng trái bắp được trồng ở các mức phân đạm khác nhau có thể được kiểm định về tính đồng nhất để xác định xem mối quan hệ tuyến tính giữa năng suất hạt và trọng lượng trái có bị ảnh hưởng bởi mức phân đạm hay không?

Chúng ta sẽ sử dụng số liệu đạm protein hòa tan (X_1) và diệp lục tố (chlorophyll) tổng số (X_2) trong lá của giống lúa IR8 và IR22 để kiểm định tính đồng nhất của k hệ số tương quan tuyến tính đơn. Số liệu của giống IR8 được trình bày ở Bảng 9.4 và số liệu của giống IR22 ở Bảng 9.5. Mục đích để xác định xem mức độ liên quan tuyến tính giữa đạm protein hòa tan và diệp lục tố tổng số trong lá có giống nhau ở hai giống lúa không? Các bước được thực hiện như sau:

Bảng 9.4: Tính hệ số tương quan tuyến tính đơn giữa đạm protein hòa tan (X_1) và diệp lục tố tổng số trong lá của giống lúa IR8 (Trích từ bảng 9.5 của Gomez and Gomez. 1984. Trang 380)

Số mẫu	N protein hòa tan, mg/lá (X_1)	Diệp lục tố tổng số mg/lá (X_2)	Bình phương của lệch số		Tích của các lệch số
			x_1^2	x_2^2	$(x_1)(x_2)$
1	0,60	0,44	0,1369	0,1444	0,1406
2	1,12	0,96	0,0225	0,0196	0,0210
3	2,10	1,90	1,2769	1,1664	1,2204
4	1,16	1,51	0,0361	0,4761	0,1311
5	0,70	0,46	0,0729	0,1296	0,0972
6	0,80	0,44	0,0289	0,1444	0,0646
7	0,32	0,04	0,4225	0,6084	0,5070
Tổng	6,80	5,75	1,9967	2,6889	2,1819
Trung bình	0,97	0,82			

Bảng 9.5: Tính hệ số tương quan tuyến tính đơn giữa đạm protein hòa tan (X_1) và diệp lục tố tổng số trong lá của giống lúa IR22.

Số mẫu	N protein hòa tan, mg/lá (X_1)	Diệp lục tố tổng số mg/lá (X_2)	Bình phương của lệch số		Tích của các lệch số
			x_1^2	x_2^2	$(x_1)(x_2)$
1	0,84	0,55	0,2704	0,4900	0,3640
2	1,24	1,24	0,0144	0,0001	0,0012
3	2,10	1,56	0,5476	0,0961	0,2294
4	2,64	2,52	1,6384	1,6129	1,6256
5	1,31	1,64	0,0025	0,1521	- 0,0195
6	1,22	1,17	0,0196	0,0064	0,0112
7	0,19	0,04	1,3689	1,4641	1,4157
Tổng	9,54	8,72	3,8618	3,8217	3,6276
Trung bình	1,36	1,25			

Bước 1: Tính hệ số tương quan tuyến tính đơn cho từng bộ số liệu. Ký hiệu các hệ số này là $r_1, r_2, \ldots, r_k$. Trong ví dụ, k = 2 hệ số tương quan tuyến tính đơn được tính là $r_1 = 0,942$ cho giống IR8 (Bảng 9.4) và $r_2 = 0,944$ cho giống IR22 (Bảng 9.5).

Bước 2: Đối với mỗi giá trị r tính ở bước 1, tính giá trị z tương ứng như sau:

$$z = 0{,}5 \ln \frac{1+r}{1-r}$$

Trong ví dụ, hai giá trị z được tính là:

$$z_1 = 0{,}5 \ln \frac{1+0{,}942}{1-0{,}942} = 1{,}756$$

$$z_2 = 0{,}5 \ln \frac{1+0{,}944}{1-0{,}944} = 1{,}774$$

Bước 3: Tính trung bình có gia trọng $\bar{z}_w$ của các giá trị z theo công thức:

$$\bar{z}_w = \frac{\sum_{i=1}^{k}(n_i - 3)z_i}{\sum_{i=1}^{k}(n_i - 3)}$$

với n_i (i = 1,..., k) là số cặp quan sát ở bộ số liệu thứ i được dùng để tính giá trị r_i ở bước 1.

$$\bar{z}_w = \frac{(7-3)(1{,}756) + (7-3)(1{,}774)}{(7-3)+(7-3)} = 1{,}765$$

Chú ý, nếu giá trị n_i giống nhau cho tất cả các bộ số liệu (như trường hợp ở ví dụ này) thì giá trị $\bar{z}_w$ đơn giản là trung bình cộng của các giá trị z_i.

Bước 4: Tính giá trị chi bình phương (χ^2) theo công thức sau:

$$\chi^2 = \sum (n_i - 3)(z_i - \bar{z}_w)^2$$

$$\chi^2 = (7-3)(1{,}756 - 1{,}765)^2 + (7-3)(1{,}774 - 1{,}765)^2 = 0{,}001$$

Bước 5: So sánh giá trị χ^2 tính với giá trị χ^2 bảng ở (k – 1) độ tự do. Loại giả thuyết đồng nhất của k hệ số tương quan tuyến tính đơn nếu giá trị χ^2 tính lớn hơn giá trị χ^2 bảng tương ứng ở mức ý nghĩa đã chọn.

Trong ví dụ, giá trị χ^2 bảng ở (k – 1) = 1 độ tự do là 3,84 ở mức ý nghĩa 5% và 6,63 ở mức ý nghĩa 1%. Vì giá trị χ^2 tính nhỏ hơn giá trị χ^2 bảng ở mức ý nghĩa 5%, nên kiểm định không ý nghĩa và không thể loại bỏ tính đồng nhất của các hệ số hồi qui tuyến tính đơn.

Bước 6: Nếu kiểm định χ^2 không ý nghĩa, đạt được giá trị hệ số tương quan tuyến tính đơn gộp ở Phụ lục XIII, dựa trên giá trị $\bar{z}_w$ được tính ở bước 3. Nếu giá trị $\bar{z}_w$ không có sẵn ở Phụ lục XIII, hệ số tương quan tuyến tính đơn gộp có thể được tính như sau:

$$r_p = \frac{e^{2\bar{z}_w} - 1}{e^{2\bar{z}_w} + 1} = \frac{e^{2(1{,}765)} - 1}{e^{2(1{,}765)} + 1} = 0{,}943$$

Do đó, các tương quan tuyến tính đơn giữa đạm protein hòa tan và diệp lục tố ở lá của giống IR8 và IR22 có thể được đo với một hệ số duy nhất là 0,943

V. TƯƠNG QUAN VÀ HỒI QUI TUYẾN TÍNH BỘI

(Multiple Linear Regression and Correlation)

Giới hạn chính của phân tích tương quan và hồi qui tuyến tính đơn là chỉ có thể áp dụng cho trường hợp có một biến độc lập. Tuy nhiên, với sự nhận thức ngày càng tăng

về sự phụ thuộc lẫn nhau giữa các nhân tố sản xuất cũng như với điều kiện sẵn có của các phương pháp thí nghiệm qua đó có thể đánh giá đồng thời nhiều nhân tố, các nhà nghiên cứu đang tăng dần việc sử dụng các thí nghiệm thừa số. Do đó, có sự gia tăng tương ứng nhu cầu để sử dụng phương pháp hồi qui mà có thể vận dụng đồng thời nhiều biến độc lập.

Phân tích hồi qui chứa nhiều hơn một biến độc lập được gọi là phân tích hồi qui bội. Khi tất cả các biến độc lập được giả định có ảnh hưởng đến biến phụ thuộc theo dạng tuyến tính và độc lập lẫn nhau, phương pháp sẽ được gọi là phân tích hồi qui tuyến tính bội. Phương pháp này biểu diễn dạng quan hệ của biến phụ thuộc Y với k biến độc lập $X_1, X_2, \ldots, X_k$ có thể được biểu diễn ở dạng hàm số:

$$Y = \alpha + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_k X_k$$

Số liệu đòi hỏi để áp dụng phân tích hồi qui tuyến tính bội có k biến độc lập là (n)(k+1) số quan sát được mô tả ở đây:

Số Quan sát	Giá trị quan sát					
	Y	X_1	X_2	X_3	...	X_k
1	Y_1	X_{11}	X_{21}	X_{31}	...	X_{k1}
2	Y_2	X_{12}	X_{22}	X_{32}	...	X_{k2}
3	Y_3	X_{13}	X_{23}	X_{33}	...	X_{k3}
.	.	.	.	.	...	.
.	.	.	.	.	...	.
.	.	.	.	.	...	.
n	Y_n	X_{1n}	X_{2n}	X_{3n}	...	X_{kn}

(k + 1) biến Y, X_1, X_2, ..., X_k phải được đánh giá đồng thời đối với một trong n đơn vị quan sát (nghĩa là, đơn vị thí nghiệm hoặc đơn vị lấy mẫu). Thêm vào đó phải có đủ số quan sát để thực hiện n lớn hơn (k + 1).

Phương pháp hồi qui tuyến tính bội bao gồm việc ước lượng và kiểm định ý nghĩa của (k + 1) tham số của phương trình hồi qui tuyến tính bội. Để giải thích phương pháp, chúng ta sẽ lấy ví dụ cho trường hợp k = 2 bằng cách sử dụng số liệu năng suất hạt (Y), chiều cao cây (X_1) và số chồi (X_2) trong Bảng 9.6. Với k = 2, phương trình hồi qui bội được biểu diễn như sau:

$$Y = \alpha + \beta_1 X_1 + \beta_2 X_2$$

Bước 1: Tính trung bình và tổng bình phương đã hiệu chỉnh cho mỗi (k + 1) biến Y, X_1, X_2, ..., X_k và tổng tích chéo đã hiệu chỉnh cho tất cả các cặp tổ hợp có thể của (k + 1) biến. Tóm tắt các tham số tính được, cùng với các biến như sau:

Biến	Trung bình	Tổng bình phương và tổng tích chéo đã hiệu chỉnh				
		X_1	X_2	...	X_k	Y
X_1	$\overline{X}_1$	$\sum x_1^2$	$\sum x_1 x_2$	...	$\sum x_1 x_k$	$\sum x_1 y$
X_2	$\overline{X}_2$		$\sum x_2^2$	...	$\sum x_2 x_k$	$\sum x_2 y$
.	.				.	.
.	.				.	.
.	.				.	.
X_k	$\overline{X}_k$				$\sum x_k^2$	$\sum x_k y$
Y	$\overline{Y}$					$\sum y^2$

Bảng 9.6: Tính phương trình hồi qui tuyến tính bội. Sự quan hệ giữa chiều cao cây (X_1) và số chồi (X_2) đến năng suất (Y) trên tám giống lúa.

Số giống	Năng suất, kg/ha (Y)	Chiều cao cây, cm (X_1)	Số chồi/bụi (X_2)
1	5755	110,5	14,5
2	5939	105,4	16,0
3	6010	118,1	14,6
4	6545	104,5	18,2
5	6730	93,6	15,4
6	6750	84,1	17,6
7	6899	77,8	17,9
8	7862	75,6	19,4
Trung bình	**6561**	**96,2**	**16,7**
$\sum x_1^2 = 1753,72$	$\sum x_2^2 = 23,22$		$\sum x_1 y = -65194$
$\sum x_2 y = 7210$	$\sum x_1 x_2 = -156,65$		$\sum y^2 = 3211504$

Ví dụ: Kết quả tính từ Bảng 9.6

Bước 2: Giải $b_1, b_2, \ldots, b_k$ từ k phương trình đồng thời sau đây mà nhìn chung liên quan đến các phương trình chuẩn:

$$b_1 \sum x_1^2 + b_2 \sum x_1 x_2 + \cdots + b_k \sum x_1 x_k = \sum x_1 y$$
$$b_1 \sum x_1 x_2 + b_2 \sum x_2^2 + \cdots + b_k \sum x_2 x_k = \sum x_2 y$$
$$\vdots$$
$$b_1 \sum x_1 x_k + b_2 \sum x_2 x_k + \cdots + b_k \sum x_k^2 = \sum x_k y$$

với $b_1, b_2, \ldots, b_k$ là các giá trị ước lượng của $\beta_1, \beta_2, \ldots, \beta_k$ của phương trình hồi qui tuyến tính bội, và các giá trị của tổng bình phương và tổng tích chéo của (k + 1) biến là các giá trị được tính ở bước 1.

Có nhiều phương pháp chuẩn tắc hóa để giải k phương trình cùng một lúc cho k không biết, hoặc bằng tay hoặc với sự trợ giúp của máy tính (Simmons, 1948, Anderson và Bancroft, 1952, Nie và ctv, 1975 và Barr và ctv., 1979).

Trong ví dụ, với k = 2, phương trình chuẩn là:

$$b_1 \sum x_1^2 + b_2 \sum x_1 x_2 = \sum x_1 y$$
$$b_1 \sum x_1 x_2 + b_2 \sum x_2^2 = \sum x_2 y$$

Và cách giải b_1 và b_2 là:

$$b_1 = \frac{\left(\sum x_2^2\right)\left(\sum x_1 y\right) - \left(\sum x_1 x_2\right)\left(\sum x_2 y\right)}{\left(\sum x_1^2\right)\left(\sum x_2^2\right) - \left(\sum x_1 x_2\right)^2}$$

$$= \frac{(23,22)(-65194) - (-156,65)(7210)}{(1753,72)(23,22) - (-156,65)^2} = -23,75$$

$$b_2 = \frac{\left(\sum x_1^2\right)\left(\sum x_2 y\right) - \left(\sum x_1 x_2\right)\left(\sum x_1 y\right)}{\left(\sum x_1^2\right)\left(\sum x_2^2\right) - \left(\sum x_1 x_2\right)^2}$$

$$= \frac{(1753,72)(7210) - (-156,65)(-65194)}{(1753,72)(23,22) - (-156,65)^2}$$

$$= 150,27$$

Bước 3: Tính giá trị ước lượng của tung độ gốc (α)

$$a = \overline{Y} - b_1\overline{X}_1 - b_2\overline{X}_2 - \cdots - b_k\overline{X}_k$$

với $\overline{Y}$, $\overline{X}_1$, $\overline{X}_2$, . . . , $\overline{X}_k$ là các trung bình của (k + 1) biến được tính ở bước 1.

Do đó, phương trình hồi qui tuyến tính bội được ước lượng như sau

$$\hat{Y} = a + b_1X_1 + b_2X_2 + \cdots + b_kX_k$$

Trong ví dụ, ước lượng của tung độ gốc được tính như sau:

$a = \overline{Y} - b_1\overline{X}_1 - b_2\overline{X}_2$

$= 6561 - (-23,75)(96,2) - (150,27)(16,7)$

$= 6336$

Vậy phương trình hồi qui tuyến tính bội được ước lượng giữa chiều cao cây (X_1) và số chồi (X_2) với năng suất (Y) là:

$\hat{Y} = 6336 - 23,75X_1 + 150,27X_2$

Bước 4: Tính

- Tổng bình phương do hồi qui

$$SSR = \sum_{i=1}^{k}(b_1)\left(\sum x_1 y\right)$$

- Tổng bình phương sai số

$$SSE = \sum y^2 - SSR$$

- Hệ số xác định

$$R^2 = \frac{SSR}{\sum y^2}$$

Hệ số xác định R^2 đo sự đóng góp của hàm tuyến tính của k biến độc lập đối với biến động ở Y. Thường nó được biểu diễn bằng phần trăm. Căn bậc hai của R^2 chính là hệ số tương quan bội.

Trong ví dụ, các giá trị SSR, R^2 và SSE được tính như sau:

$$SSR = \sum_{i=1}^{k}(b_1)\left(\sum x_1 y\right)$$

$= (-23,75)(-65194) + (150,27)(7210)$

$= 2631804$

$SSE = \sum y^2 - SSR = 3211504 - 2631804$

$= 579700$

$$R^2 = \frac{SSR}{\sum y^2} = \frac{2631,804}{3211504} = 0,82$$

Do đó, 82% biến động tổng cộng ở năng suất của tám giống lúa có thể được giải thích bằng hàm tuyến tính, gồm chiều cao cây và số chồi như được biểu diễn ở bước 1.

Bước 5: Kiểm định ý nghĩa R^2

Tính giá trị F như sau:

$$F = \frac{SSR / k}{SSE / (n - k - 1)}$$

Trong ví dụ, giá trị F

$$F = \frac{2631,804/2}{579700/(8-2-1)}$$

$$= 11,35$$

- So sánh giá trị F tính với giá trị F bảng với $df_1 = k$ và $df_2 = (n - k - 1)$ độ tự do. Hệ số xác định R^2 được xem như có ý nghĩa ($\neq 0$ có ý nghĩa) nếu giá trị F tính lớn hơn giá trị F bảng tương ứng ở mức ý nghĩa đã chọn.

Trong ví dụ, giá trị F bảng với $df_1 = 2$ và $df_2 = 5$ độ tự do là 5,79 ở mức ý nghĩa 5% và 13,27 ở mức 1%. Vì giá trị F tính lớn hơn giá trị F bảng ở mức ý nghĩa 5%, nhưng nhỏ hơn giá trị F bảng ở mức ý nghĩa 1%, nên phương trình hồi qui tuyến tính được ước lượng $\hat{Y} = 6336 - 23,75X_1 + 150,27X_2$ có ý nghĩa ở mức ý nghĩa 5%. Do đó, ảnh hưởng tuyến tính phối hợp của chiều cao cây và số chồi đóng góp đáng kể đến biến động năng suất.

Điều cần nhấn mạnh ở đây có liên quan đến kiểm định F có ý nghĩa (chứng tỏ R^2 có ý nghĩa) và độ lớn của giá trị R^2. Mặc dù hồi qui tuyến tính có ý nghĩa hàm ý rằng phần biến động nào đó ở Y được giải thích thật sự bởi hàm tuyến tính của các biến độc lập, độ lớn của giá trị R^2 cung cấp thông tin về độ lớn của phần đó. Dĩ nhiên, giá trị R^2 càng lớn, tầm quan trọng của phương trình hồi qui mô tả Y càng cao. Mặt khác, nếu giá trị R^2 thấp, ngay cả nếu kiểm định F có ý nghĩa, phương trình hồi qui ước lượng không thể có ý nghĩa. Ví dụ, giá trị R^2 là 0,26, ngay cả có ý nghĩa, chứng tỏ rằng chỉ có 26% biến động tổng cộng ở biến phụ thuộc Y được giải thích bởi hàm tuyến tính của các biến độc lập có liên quan. Nói cách khác, 74% biến động ở Y không thể được giải thích bởi hồi qui. Với mức ảnh hưởng thấp như thế, phương trình hồi qui ước lượng sẽ không có ích trong việc ước lượng, dự đoán ít nhiều, các giá trị của Y.

Sau đây là các gợi ý quan trọng để áp dụng phương pháp hồi qui tuyến tính bội vừa mô tả:

- Phương pháp chỉ có thể áp dụng nếu hai điều kiện sau đây thỏa. Đầu tiên, ảnh hưởng của mỗi và tất cả k biến độc lập $X_1, X_2, \ldots, X_k$ trên biến phụ thuộc Y phải tuyến tính. Nghĩa là, số lượng thay đổi ở Y trên đơn vị thay đổi ở mỗi X_i là hằng số từ đầu đến cuối dãy giá trị X_i ở điều kiện nghiên cứu. Thứ hai, ảnh hưởng của mỗi X_i trên Y thì độc lập với biến X khác. Nghĩa là, số lượng thay đổi ở Y trên đơn vị thay đổi ở mỗi X_i thì giống nhau bất chấp các giá trị của biến X khác. Bất cứ lúc nào một hoặc cả hai điều kiện nêu trên bị nới lỏng, nên sử dụng phương pháp hồi qui phi tuyến.

- Phương pháp để ước lượng và kiểm định ý nghĩa của (k + 1) tham số trong hồi qui tuyến tính bội có k biến độc lập thì khá dài và mất nhiều thời gian. Sự phức tạp càng gia tăng khi k càng lớn.

B. QUAN HỆ PHI TUYẾN *(Nonlinear Relationship)*

Quan hệ hàm số giữa hai biến được xem là phi tuyến nếu mức độ thay đổi ở Y liên quan đến một đơn vị thay đổi ở X không phải là hằng số trên dãy giá trị X định rõ. Quan hệ phi tuyến giữa các biến cũng khá phổ biến trong sinh học, đặc biệt nếu dãy giá trị rộng. Hai ví dụ điển Hình là:

- Sự đáp ứng của năng suất lúa đối với việc bón phân đạm. Sự đáp ứng này thường nhanh ở mức độ thấp, chậm hơn ở mức độ trung bình và có thể trở nên âm ở mức độ đạm cao.

- Kiểu tăng trưởng của cây theo thời gian: thường bắt đầu chậm, gia tăng nhanh ở các giai đoạn tăng trưởng trung gian và chậm lại vào giai đoạn cuối của chu kỳ sống.

Khi sự quan hệ giữa các biến ở điều kiện khảo sát không tuyến tính. Phương pháp phân tích hồi qui tuyến tính không thoả đáng, nhà nghiên cứu phải chuyển sang phương pháp phân tích hồi qui phi tuyến.

I. HỒI QUI PHI TUYẾN ĐƠN *(Simple Nonlinear Regression)*

Có nhiều dạng hàm số có thể mô tả quan hệ phi tuyến giữa hai biến, việc chọn phương pháp tương quan và hồi qui thích hợp tùy thuộc vào dạng hàm số có liên quan. Chúng ta tập trung đầu tiên vào một kỹ thuật: mà bao hàm việc tuyến tính hóa dạng phi tuyến hoặc qua việc chuyển đổi biến hoặc qua việc tạo các biến mới. Chúng ta chỉ tập trung vào một kỹ thuật với hai lý do:

- Kỹ thuật đơn giản, bởi vì sau khi tuyến tính hóa có thể áp dụng phương pháp hồi qui trực tiếp.

- Kỹ thuật có thể áp dụng rộng rãi, vì hầu hết các quan hệ phi tuyến được tìm thấy trong nghiên cứu nông nghiệp có thể được tuyến tính hóa qua việc chuyển đổi biến hoặc tạo biến mới.

1. Chuyển đổi biến (Transformation of Variable).

Quan hệ phi tuyến giữa hai biến có thể được tuyến tính hóa bằng cách chuyển đổi một hoặc cả hai biến. Dưới đây là các ví dụ của một vài dạng phi tuyến, được khuyến cáo phổ biến trong nông nghiệp mà có thể tuyến tính hóa qua việc chuyển đổi biến.

Ví dụ 1: Dạng phi tuyến

$$Y = ae^{\beta X}$$

có thể được tuyến tính hóa bằng cách chuyển đổi biến phụ thuộc Y sang lnY, với ln chỉ logarith Néper (log cơ số e). Do đó, dạng tuyến hóa là

$$Y = \alpha' + \beta X$$

với $Y' = \ln Y$ và $\alpha' = \ln \alpha$

Ví dụ 2: Dạng phi tuyến

$$Y = \alpha\beta^X$$

có thể được tuyến tính hóa bằng cách chuyển đổi biến phụ thuộc Y sang logY, với log chỉ logarith cơ số 10, Do đó, dạng tuyến hóa là

$$Y = \alpha' + \beta' X$$

với $Y = \log Y$, $\alpha' = \log \alpha$ và $\beta' = \log \beta$.

Ví dụ 3: Dạng tuyến tính

$$\frac{1}{Y} = \alpha + \beta X$$

có thể được tuyến tính hóa bằng cách chuyển đổi biến phụ thuộc Y sang 1/Y. Do đó, dạng tuyến tính hóa là

$$Y' = \alpha + \beta X$$

với $Y' = 1/Y$

Ví dụ 4: Dạng tuyến tính

$$Y = \alpha + \frac{\beta}{X}$$

có thể được tuyến tính hóa bằng cách chuyển đổi biến độc lập X thành 1/X. Do đó, dạng tuyến tính hóa là

$$Y = \alpha + \beta X''$$

với X ' = 1/X

Ví dụ 5: Dạng phi tuyến

$$Y = \left(\alpha + \frac{\beta}{X}\right)^{-1}$$

có thể được tuyến tính hóa bằng cách chuyển đổi biến phụ thuộc Y thành 1/Y và biến độc lập X thành 1/X. Do đó, dạng tuyến tính hóa là

$$Y' = \alpha + \beta X''$$

với Y' = 1/Y và X ' = 1/X

Sau khi tuyến tính hóa qua cách chuyển đổi biến, phương pháp tương quan và hồi qui tuyến tính đơn có thể được áp dụng trực tiếp. Ví dụ sau đây là phương pháp chuyển đổi sang dạng hồi qui tuyến tính đơn với số liệu về tỷ lệ truyền ánh sáng (Y) và chỉ số diện tích lá (X) trong Bảng 9.7 phù hợp với dạng phi tuyến của ví dụ 1, nghĩa là

$$Y = ae^{\beta X}$$

Bảng 9.7: Tính phương trình hồi qui phi tuyến giữa tỷ lệ truyền ánh sáng (Y) và chỉ số diện tích lá (X) của giống lúa IR8 bằng phương pháp chuyển đổi biến (Trích từ bảng 9.7 của Gomez and gomez. 1984. Trang 390)

Số quan sát	Tỷ lệ truyền Ánh sáng (Y)	Chỉ số diện tích lá (X)	Y' = lnY
1	75,0	0,50	4,31749
2	72,0	0,60	4,27667
3	42,0	1,80	3,73767
4	29,0	2,50	3,36730
5	27,0	2,80	3,29584
6	10,0	5,45	2,30259
7	9,0	5,60	2,19722
8	5,0	7,20	1,60944
9	2,0	8,75	0,69315
10	2,0	9,60	0,69315
11	1,0	10,40	0,00000
12	0,9	12,00	- 0,10536
Trung bình	22,9	5,60	- 2,19876
$\sum y'^2 = 28,77936$	$\sum x^2 = 175,40500$	$\sum xy' = -70,76124$	

Các bước thực hiện:

Bước 1: Tuyến tính hóa dạng hàm phi tuyến quy định qua việc biến đổi thích hợp một hoặc hai biến. Với ví dụ, dạng tuyến tính hóa của hàm số nêu trên là:

$$Y = \alpha' + \beta X$$

với $Y' = \ln Y$ và $\alpha' = \ln \alpha$

Bước 2: Tính các giá trị đã biến đổi cho tất cả n đơn vị quan sát của mỗi biến đã được biến đổi ở bước 1. Trong ví dụ, các giá trị $Y' = \ln Y$ được tính cho n = 12 cặp giá trị (X, Y) được trình bày trong Bảng 9.7.

Bước 3: Áp dụng kỹ thuật hồi qui tuyến tính đơn đối với số liệu nhận được ở bước 2 dựa vào dạng tuyến tính lấy ở bước 1. Trong ví dụ, các ước lượng của hai tham số α' và β được tính như sau:

- Tính các trung bình $\overline{Y'}$ và $\overline{X}$, tổng bình phương $\sum y'^2$, $\sum x^2$ và tổng tích chéo $\sum xy'$ của hai biến Y' và X như ghi ở Bảng 9.7.

- Tính các giá trị ước lượng của α' và β như sau:

$$b = \frac{\sum xy'}{\sum x^2} = \frac{-70,76124}{175,40500} = -0,40342$$

$$\alpha' = \overline{Y'} - b\overline{X}$$
$$= 2,19876 - (-0,40342)(5,60)$$
$$= 4,45791$$

Bước 4: Sử dụng giá trị ước lượng của các tham số hồi qui dạng tuyến tính hóa đạt được ở bước 3, lấy giá trị ước lượng thích hợp của hồi qui ban đầu dựa vào cách biến đổi riêng biệt được sử dụng ở bước 1.

Trong ví dụ, cần hoán chuyển các tham số hồi qui ước lượng. Nghĩa là, để lấy giá trị ước lượng của hồi qui phi tuyến ban đầu, tham số hồi qui α cần được tính bằng antilog của α':

α = antilog của α' = antilog của 4,45791
= 86,31

và dùng ước lượng của β tính ở bước 3, ước lượng phương trình hồi qui yêu cầu đạt được như sau:

$$\hat{Y} = 86,31e^{-0,40342X}$$

Trình bày bằng đồ thị hồi qui phi tuyến ước lượng như chỉ ở Hình 9.12.

2. Tạo biến mới *(Creation of New variable)*

Quan hệ phi tuyến giữa hai biến có thể được tuyến tính hóa qua việc tạo ra một hoặc nhiều biến mới để chúng có thể giải thích thành phần phi tuyến của hàm ban đầu. Trong nghiên cứu nông nghiệp, kỹ thuật này được áp dụng phổ biến nhất đối với đa thức mức bậc k:

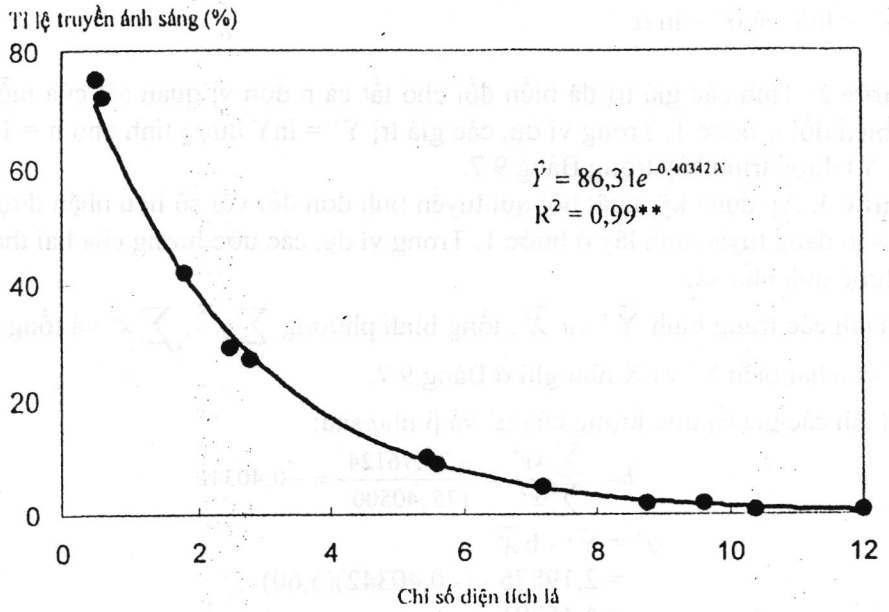
$$Y = \alpha + \beta_1 X + \beta_2 X^2 + \cdots + \beta_k X^k$$

với $Z_1 = X$, $Z_2 = X^2$, . . . , và $Z_k = X^k$

Với dạng tuyến hóa đến từ việc tạo biến mới, có thể áp dụng phương pháp phân tích tương quan và hồi qui tuyến tính bội.

Sử dụng số liệu của Bảng 9.8 để giải thích kỹ thuật này, ở đó quan hệ giữa năng suất (Y) và mức độ đạm (X) được giả định là bậc hai (đa thức bậc 2).

$$Y = \alpha + \beta_1 X + \beta_2 X^2$$



Hình 9.12: Quan hệ phi tuyến được ước lượng giữa tỷ lệ truyền ánh sáng (Y) và chỉ số diện tích lá (X) của giống lúa IR8.

Các bước thực hiện như sau:

Bước 1: Tuyến tính hóa dạng hàm phi tuyến được qui định qua việc tạo bộ biến mới thích hợp.

Trong ví dụ, dạng tuyến tính hóa của đa thức bậc hai là

$$Y = \alpha + \beta_1 Z_1 + \beta_2 Z_2$$

với hai biến mới tạo được định nghĩa $Z_1 = X$ và $Z_2 = X^2$

Bước 2: Tính giá trị của mỗi biến mới được tạo cho tất cả n đơn vị quan sát. Trong ví dụ, chỉ có các giá trị của biến Z_2 cần được tính vì các giá trị của biến Z_1 giống với giá trị của biến X ban đầu. Các giá trị Z_2 được tính bằng bình phương các giá trị tương ứng của biến X ban đầu, như chỉ ở Bảng 9.8.

Bước 3: Áp dụng kỹ thuật hồi qui tuyến tính bội thích hợp đối với dạng tuyến tính hóa nhận được ở bước 1, bằng cách sử dụng số liệu lấy ở bước 2. Trong ví dụ, vì dạng tuyến tính hóa gồm hai biến độc lập Z_1 và Z_2, phương pháp hồi qui tuyến tính bội đối với hai biến độc lập được áp dụng như sau:

- Tính trung bình, tổng bình phương và tổng tích chéo cho ba biến Y, Z_1 và Z_1 như chỉ ở bảng.

- Tính ước lượng của ba tham số: α, β_1 và β_2, theo công thức ở bước 2:

$$b_1 = \frac{\left(\sum z_2^2\right)\left(\sum z_1 y\right) - \left(\sum z_1 z_2\right)\left(\sum z_2 y\right)}{\left(\sum z_1^2\right)\left(\sum z_2^2\right) - \left(\sum z_1 z_2\right)^2}$$

$$= \frac{(140940000)(112530) - (1080000)(12167100)}{(9000)(140940000) - (1080000)^2}$$

$$= 26,65$$

$$b_2 = \frac{\left(\sum z_1^2\right)\left(\sum z_2 y\right) - \left(\sum z_1 z_2\right)\left(\sum z_1 y\right)}{\left(\sum z_1^2\right)\left(\sum z_2^2\right) - \left(\sum z_1 z_2\right)^2}$$

$$= \frac{(9000)(12167100) - (1080000)(112530)}{(9000)(140940000) - (1080000)^2}$$

$$= -0{,}118$$

$$a = \bar{Y} - b_1\bar{Z}_1 - b_2\bar{Z}_2$$

$$= 5824 - (26{,}65)(60) - (-0{,}118)(5400)$$

$$= 4862$$

Bảng 9.8: Tính phương trình hồi qui bậc hai ước lượng, hoặc dạng tuyến tính hóa: để mô tả phản ứng năng suất của giống lúa IR661-1-170 đối với phân đạm (Trích từ bảng 9.8 của Gomez and gomez. 1984. Trang 393).

Số cặp	Năng suất hạt Kg/ha (X)	Mức độ đạm Kg/ha ($Z_1 = X$)	$Z_2 = X^2$
1	4,878	0	0
2	5,506	30	900
3	6,083	60	3600
4	6,291	90	8100
5	6,361	120	14400
Trung bình	5,824	60	5400
$\sum y^2 = 1569579$	$\sum z_1^2 = 9000$		$\sum z_2^2 = 140940000$
$\sum z_1 y = 112{,}530$	$\sum z_2 y = 12167100$		$\sum z_2 z_2 = 1080000$

Do đó, phương trình hồi qui đa thức bậc hai mô tả phản ứng năng suất của giống lúa IR661-1-170 đối với các mức độ đạm được áp dụng trong khoảng 0 đến 120 kg/ha được ước lượng như sau:

$$\hat{Y} = 4862 + 26{,}65X - 0{,}118X^2 \quad \text{với } 0 \le X \le 120$$

- Tính hệ số xác định như sau:

$$R^2 = \frac{b_1\sum z_1 y + b_2\sum z_2 y}{\sum y^2}$$

$$= \frac{(26{,}65)(112530) + (-0{,}118)(12167100)}{1569579}$$

$$= \frac{1563207}{1569579} = 0{,}996$$

- Tính giá trị F:

$$F = \frac{(n-k-1)(b_1\sum z_1 y + b_2\sum z_2 y)}{k(\sum y^2 - b_1\sum z_1 y - b_2\sum z_2 y)}$$

$$= \frac{(5-2-1)(1563207\ \ )}{(2)(156957\ \ 9 - 1563207)}$$

$$= 245{,}32$$

Vì giá trị F tính vượt giá trị F bảng với $df_1 = df_2 = 2$ độ tự do ở mức ý nghĩa 1% là 99,0, nên phương trình hồi qui bậc 2 ước lượng có ý nghĩa ở mức 1%. Kết quả chỉ rằng phản ứng năng suất của giống lúa IR661-1-170 đối với phân đạm có thể được mô tả

dầy đủ bằng phương trình bậc hai. Giá trị R^2 là 0,996 chứng tỏ rằng 99,6% biến động tổng cộng ở năng suất trung bình được giải thích bởi phương trình hồi qui bậc 2 được ước lượng.

II. HỒI QUI PHI TUYẾN BỘI *(Multiple Nonlinear Regression)*

Khi sự quan hệ giữa biến phụ thuộc Y và k biến độc lập $X_1, X_2, \ldots, X_k$, với k > 1, không tuân theo quan hệ tuyến tính bội, lúc đó tồn tại quan hệ phi tuyến bội. Sự xuất hiện của quan hệ phi tuyến bội có thể là kết quả của bất kỳ trường hợp sau:

- Ít nhất một trong các biến độc lập biểu hiện quan hệ phi tuyến với biến phụ thuộc Y. Ví dụ, với hai biến độc lập X_1 và X_2, quan hệ phi tuyến bội tồn tại nếu một hoặc cả hai biến biểu thị quan hệ phi tuyến với biến phụ thuộc. Nếu cả hai X_1, X_2 có quan hệ với Y theo dạng bậc hai chẳng hạn, phương trình hồi qui phi tuyến bội tương ứng trình bày sự quan hệ của chúng đối với Y sẽ là:

$$Y = \alpha + \beta_1 X_1 + \beta_2 X_1^2 + \beta_3 X^2 + \beta_4 X_2^2$$

- Ít nhất hai biến độc lập tương tác lẫn nhau. Ví dụ, có hai biến độc lập X_1 và X_2, một biến ảnh hưởng đến Y ở dạng tuyến tính, phương trình hồi qui bội có lẽ là phi tuyến nếu tác động của nhân tố X_1 trên Y thay đổi theo mức độ của nhân tố X_2 và ngược lại. Trong trường hợp như thế, phương trình hồi qui phi tuyến bội có thể được tiêu biểu bởi:

$$Y = \alpha + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_1 X_2$$

với đại lượng sau cùng của phương trình trình bày dạng tương tác.

- Cả hai trường hợp nêu trên xảy ra đồng thời. Nghĩa là, ít nhất một trong các biến độc lập có quan hệ phi tuyến với biến phụ thuộc và ít nhất hai biến độc lập tương tác lẫn nhau. Đặt hai phương trình cùng với nhau, chúng có thể có:

$$Y = \alpha + \beta_1 X_1 + \beta_2 X_1^2 + \beta_3 X_2 + \beta_4 X_2^2 + \beta_5 X_1 X_2$$

hoặc tổng quát:

$$Y = \alpha + \beta_1 X_1 + \beta_2 X_1^2 + \beta_3 X_2 + \beta_4 X_2^2 + \beta_5 X_1 X_2 + \beta_5 X_1^2 X_2 + \beta_7 X_1 X_1^2 + \beta_8 X_1^2 X_2^2$$

Ba ví dụ của dạng phi tuyến bội và dạng tuyến tính hóa của chúng là:

Ví dụ 1: Tuyến tính hóa của phương trình

$$Y = \alpha + \beta_1 X_1 + \beta_2 X_1^2 + \beta_3 X_2 + \beta_4 X_2^2$$

qua việc tạo hai biến mới $Z = X_1^2$ và $Z = X_2^2$. Do đó dạng tuyến tính hóa là:

$$Y = \alpha + \beta_1 X_1 + \beta_2 Z_2 + \beta_3 X_2 + \beta_4 Z_2$$

Ví dụ 2: Tuyến tính hóa phương trình Cobb-Douglas:

$$Y = \alpha \beta_1^{X_1} . \beta_2^{X_2} . \beta_3^{X_3} \ldots \beta_k^{X_k}$$

qua việc chuyển đổi biến phụ thuộc Y sang logY. Do đó, dạng tuyến tính hóa là:

$$Y' = \alpha' + \beta_1' X_1 + \beta_2' X_2 + \cdots + \beta_k' X_k$$

với $Y' = \log Y$, $\alpha' = \log \alpha$ và $\beta_i' = \log \beta_i$ (I = 1,..., k)

Ví dụ 3: Tuyến tính hóa dạng phi tuyến, gây ra bởi sự hiện diện của một hoặc nhiều số hạng tương tác, qua việc tạo một biến mới cho mỗi số hạng tương tác. Ví dụ, với phương trình sau đây có một số hạng tương tác.

$$Y = \alpha + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_1 X_2$$

chỉ có một biến mới $Z_1 = X_1 X_2$ cần được tạo ra. Dạng tuyến tính hóa lúc đó là:

$$Y = \alpha + \beta_1 X_1 + \beta_2 X_2 + \beta_3 Z_1$$

Một lần rút ra dạng tuyến tính hóa, có thể áp dụng trực tiếp phương pháp phân tích hồi qui tuyến tính bội .

CHƯƠNG 10

KIỂM ĐỊNH CHI BÌNH PHƯƠNG

Kiểm định chi bình phương được sử dụng phổ biến nhất để kiểm định giả thuyết liên quan đến phân bố tần số của một hoặc nhiều tổng thể. Chúng ta tập trung vào bốn cách sử dụng kiểm định chi bình phương phổ biến nhất trong nghiên cứu nông nghiệp: Phân tích số liệu thuộc tính, kiểm định cho tính đồng nhất của phương sai, kiểm định cho độ lệch chuẩn của tổng thể và kiểm định cho tính phù hợp.

I. PHÂN TÍCH SỐ LIỆU THUỘC TÍNH

Số liệu thu thập từ các thí nghiệm về nông nghiệp có thể là số liệu đo lường hoặc số liệu thuộc tính. Số liệu đo lường được ghi nhận theo tỷ lệ bằng số liên tục như năng suất, chiều cao cây, hàm lượng protein,...; nhưng số liệu thuộc tính có liên quan đến con số hữu hạn của các nhóm rời rạc. Loại số liệu thuộc tính phổ biến nhất là loại có hai nhóm, bao gồm sự hiện diện hoặc vắng mặt của một thuộc tính như đực hoặc cái, thành công hoặc thất bại, hiệu quả hoặc không hiệu quả,... Loại số liệu thuộc tính có nhiều hơn hai nhóm chẳng hạn như sự phân nhóm giống, phân nhóm màu sắc và tình trạng chiếm hữu của nông dân.

Số nhóm rời rạc trong số liệu thuộc tính có thể được ghi nhận dựa trên một hoặc nhiều tiêu chuẩn phân nhóm. Khi chỉ sử dụng một tiêu chuẩn phân nhóm, số liệu thuộc tính có liên quan đến sự phân nhóm một chiều *(One-way classification)*, chẳng hạn sự hiện diện hoặc vắng mặt của một đặc tính, sự phân nhóm màu sắc của một mô thực vật, tình trạng chiếm hữu của nông dân. Khi sử dụng nhiều hơn một tiêu chuẩn phân nhóm, số liệu thuộc tính có thể liên quan đến sự phân nhóm hai chiều, phân nhóm ba chiều,... tùy thuộc vào số tiêu chuẩn phân nhóm được sử dụng. Số liệu thuộc tính có phân nhóm hai chiều tạo thành sự phân nhóm hai chiều r x c hoặc bảng liên hợp r x c; với r và c biểu thị số nhóm trong hai tiêu chuẩn phân nhóm được sử dụng. Ví dụ, nếu các giống lúa trong trắc nghiệm giống được phân nhóm dựa trên hai tiêu chuẩn – màu sắc của phiến lá (xanh hoặc tím) và loại giống (indica, japonica hoặc giống lai) – kết quả số liệu thuộc tính trình bày bảng liên hợp 2 x 3, Chú ý rằng bảng liên hợp có thể phát triển thành ba chiều, bốn chiều,...với sự thêm vào liên tục nhiều tiêu chuẩn phân nhóm.

Nhìn chung, số liệu thuộc tính đạt được khi nó không thể sử dụng số liệu đo lường. Tuy nhiên, trong một vài trường hợp đặc biệt các vật liệu thí nghiệm có thể được phân thành các nhóm rời rạc mặc dù có thể dùng được số đo định lượng. Ví dụ, cây trồng có thể phân thành ba nhóm chiều cao rời rạc (cao, trung gian hoặc thấp) thay vì được đo bằng cm. Hoặc tính kháng dọc đối với dịch sâu hại có thể được đánh giá theo cấp thang điểm từ 0 – 9 thay vì đánh giá phần trăm cây bị thiệt hại.

Có ba cách áp dụng quan trọng của kiểm định chi bình phương trong phân tích số liệu thuộc tính là:

* Kiểm định cho giả thuyết có tỷ lệ cố định
* Kiểm định cho tính độc lập trong bảng liên hợp
* Kiểm định tính đồng nhất của tỷ lệ

1. Kiểm định cho giả thuyết có tỷ lệ cố định *(Test for a fixed-ratio hypothesis)*

Đây là phương pháp kiểm định để quyết định bộ số liệu thuộc tính có phù hợp với phân bố tần số được giả thuyết dựa trên một vài hiện tượng sinh học. Kiểm định cho giả thuyết có tỷ lệ cố định có thể áp dụng cho số nhóm bất kỳ nhận được từ số tiêu chuẩn phân nhóm bất kỳ. Công thức tính giá trị χ^2 sẽ tùy thuộc vào số nhóm.

a. Trường hợp có hai nhóm

Giá trị χ^2 được tính theo công thức hiệu chỉnh của Yates như sau:

$$\chi^2 = \sum_{i=1}^{k} \frac{(|o_i - e_i| - 0,5)^2}{e_i}$$

với k là số nhóm, o_i là số quan sát của nhóm thứ i và e_i ($e_i = np_i$) là số mong đợi của nhóm thứ i.

Ví dụ: Ở lúa, người ta nghi ngờ rầy xanh có sự lựa chọn thức ăn giữa cây bệnh và cây mạnh. Một nhà nghiên cứu thả một số rầy xanh vào trong lồng có trồng các cây lúa mạnh và cây bệnh. Sau 2 giờ đặt trong lồng, ông đếm số côn trùng tìm thấy trên cây mạnh và cây bệnh. Trong số 239 rầy xanh được thả, ông thấy có 67 con trên cây mạnh và 172 con trên cây bệnh. Ông muốn biết tỷ lệ 67 : 172 có lệch đáng kể với tỷ lệ không có sự lựa chọn, nghĩa là, tỷ lệ 1 : 1 hay không?

Ông đặt các giả thuyết:

H_0: $p_1 = 1/2$, $p_2 = 1/2$ đối với H_a: $p_1 \neq 1/2$ hoặc $p_2 \neq 1/2$

Tính số mong đợi ở mỗi nhóm:

$e_1 = e_2 = np_1 = np_2 = 239(1/2) = 119,5$

Lập bảng để tính χ^2 như sau:

Nhóm	Quan sát (o_i)	Mong đợi (e_i)	$(\|o_i - e_i\| - 0,5)$	$(\|o_i - e_i\| - 0,5)^2$	$\frac{(\|o_i - e_i\| - 0,5)^2}{e_i}$
Cây mạnh	67	119,5	52	2704	22,63
Cây bệnh	172	119,5	52	2704	22,63
Tổng	239	239			45,26 = χ^2

Trong ví dụ chỉ có hai nhóm kết quả, nên độ tự do = k - 1 = 1. Tìm trong Phụ lục V, giá trị χ^2 là 3,84 ở mức ý nghĩa 5% và 6,63 ở mức ý nghĩa 1%. Vì giá trị χ^2 tính vượt quá giá trị χ^2 bảng ở mức ý nghĩa 1%, nên giả thuyết H_0 (nghĩa là, tỷ lệ 1 : 1) bị bác bỏ ở mức ý nghĩa 1%. Do đó, ông kết luận tỷ lệ quan sát khác với tỷ lệ giả thuyết (1 : 1). Mặt khác, quan sát số liệu ông nhận thấy có 72% rầy xanh xuất hiện trên cây bệnh và 28% trên cây mạnh; vì thế, ông kết luận rầy xanh thích ăn cây lúa bị bệnh hơn cây lúa mạnh.

b. Trường hợp có nhiều hơn hai nhóm

Giá trị χ^2 được tính theo công thức như sau:

$$\chi^2 = \sum_{i=1}^{k} \frac{(o_i - e_i)^2}{e_i}$$

Ví dụ: Một nhà chọn giống nghiên cứu tổ hợp lai giữa hai dòng tự phối: bắp ngọt, hạt vàng và bắp đá, hạt trắng. Ông muốn biết loại hạt và màu hạt ở quần thể F_2 có tuân theo tỷ lệ 9:3:3:1 của phép lai hai tính không? Từ cây F_1, ông thu hạt F_2 và phân loại chúng thành bốn nhóm theo màu hạt (vàng hoặc trắng) và loại hạt (đá hoặc ngọt) như sau: vàng-đá : vàng-ngọt : trắng-đá : trắng-ngọt. Giả sử ông quan sát trên 800 hạt F_2 và

ghi nhận có 496 vàng-đá, 158 vàng-ngọt, 112 trắng-đá và 34 trắng-ngọt. Lúc đó, ông đặt câu hỏi tỷ lệ quan sát 496:158:112:34 có lệch đáng kể so với tỷ lệ giả thuyết 9:3:3: 1?

Trong trường hợp này có bốn nhóm kết quả được tìm thấy từ số liệu phân nhóm hai chiều, nên độ tự do = k - 1 = 3.

Ông đặt các giả thuyết:

H_0: $p_1 = 9/16$, $p_2 = 3/16$, $p_3 = 3/16$, $p_4 = 1/16$

H_a: $p_1 \neq 9/16$ hoặc $p_2 \neq 3/16$ hoặc $p_3 \neq 3/16$ hoặc $p_4 \neq 1/16$

Nghĩa là, ít nhất có một nhóm khác với xác suất trên. Giả thuyết H_0 sẽ bị loại nếu một trong bốn kết quả thí nghiệm khác với giả thuyết H_0.

Tính số mong đợi ở mỗi nhóm

$e_1 = np_1 = (800)(9/16) = 450$

$e_2 = np_2 = (800)(3/16) = 150$

$e_3 = np_3 = (800)(3/16) = 150$

$e_4 = np_4 = (800)(1/16) = 50$

Lập bảng để tính χ^2 như sau:

	(Quan sát)	(Mong đợi)			
Nhóm	o_i	e_i	$o_i - e_i$	$(o_i - e_i)^2$	$(o_i - e_i)^2 / e_i$
Vàng - đá	496	450	46	2116	4,702
Vàng - ngọt	158	150	8	64	0,427
Trắng - đá	112	150	-38	1444	9,627
Trắng - ngọt	34	50	-16	256	5,120
Tổng	800	800			19,876

Trong Phụ lục V, giá trị χ^2 với độ tự do bằng 3 là 7,81 ở mức ý nghĩa 5% và 11,3 ở mức ý nghĩa 1%. Trong ví dụ, vì giá trị χ^2 tính lớn hơn 11,3 nên nhà chọn giống kết luận tỷ lệ quan sát khác với tỷ lệ lý thuyết 9:3:3:1.

2. Kiểm định cho tính độc lập trong bảng liên hợp

(Test for independence in a contingency table)

Khi số nhóm trong số liệu thuộc tính được dựa trên hai tiêu chuẩn phân nhóm, một có r nhóm và một tiêu chuẩn khác có c nhóm, kết quả số liệu tạo thành bảng liên hợp r x c. Ví dụ, một nhà nghiên cứu muốn biết café có gây ra hiện tượng mất ngũ hay không? Với hai nhóm phân biệt của hành động (tiêu chuẩn phân nhóm thứ nhất): dùng và không dùng café; và hai nhóm tình trạng (tiêu chuẩn phân nhóm thứ hai): mất ngũ và không mất ngũ. Kết quả số liệu tạo thành bảng liên hợp 2 x 2. Đối với số liệu như thế, mỗi người được lấy mẫu có thể được xếp vào một trong bốn nhóm: dùng café và mất ngũ, dùng café và không mất ngũ, không dùng café và mất ngũ, và không dùng café và không mất ngũ.

Với bảng liên hợp như thế, câu hỏi thường được đặt ra là tỷ lệ của các nhóm khác nhau trong tiêu chuẩn phân nhóm thứ nhất vẫn giống nhau trên tất cả các nhóm của tiêu chuẩn phân nhóm thứ hai và ngược lại. Nếu trả lời có, hai tiêu chuẩn phân nhóm được xem là độc lập.

Trong ví dụ, câu hỏi được đặt ra là có phải tỷ lệ người bị mất ngũ và không mất ngũ vẫn giống nhau cho cả hai nhóm: dùng và không dùng café hoặc có phải tỷ lệ người bị mất ngũ và không mất ngũ độc lập với hành động dùng và không dùng café? Kiểm định χ^2 cho tính độc lập trong bảng liên hợp là phương pháp thích hợp để trả lời câu hỏi này. Giá trị χ^2 dùng để kiểm định tính độc lập được tính theo công thức:

$$\chi^2 = \frac{\sum (o_{ij} - e_{ij})^2}{e_{ij}}$$

với o_{ij} là giá trị quan sát của hàng thứ i và cột thứ j

e_{ij} là giá trị mong đợi của hàng thứ i và cột thứ j và được tính như sau:

$$e_{ij} = \frac{R_i C_j}{N}$$

Trường hợp có hai tiêu chuẩn phân nhóm, với mỗi tiêu chuẩn bao gồm hai nhóm tạo thành bảng liên hợp 2 x 2, giá trị χ^2 được tính theo công thức hiệu chỉnh của Yates.

$$\chi^2 = \frac{\left(\sum_{i,j}^{r,c} |o_{ij} - e_{ij}| - 0,5\right)^2}{e_{ij}}$$

Tuy nhiên, nếu sử dụng các công thức trên, chúng ta cần phải tính các tần số mong đợi (e_{ij}). Do đó, trong thực hành người ta thường dùng các công thức đơn giản hơn để tính χ^2.

a. Trường hợp bảng liên hợp 2 x 2

Đặc tính 1 / Đặc tính 2	B_1	B_2	Tổng hàng (R)
A_1	a	b	R_1
A_2	c	d	R_2
Tổng cột (C)	C_1	C_2	N

Giá trị χ^2 được tính theo công thức như sau:

$$\chi^2 = \frac{N\left(|ad - bc| - \frac{N}{2}\right)^2}{R_1 R_2 C_1 C_2}$$

Lấy lại ví dụ café có gây ra hiện tượng mất ngũ hay không? Một nhà nghiên cứu điều tra 168 người và ghi nhận được kết quả như sau:

Tình trạng / Hành động	Mất ngũ	Không mất ngũ	Tổng hàng (R)
Dùng café	30	48	78
Không dùng café	15	75	90
Tổng cột (C)	45	123	168

Các bước thực hiện kiểm định:

Bước 1: Đặt các giả thuyết:

H_0: Café không gây ra hiện tượng mất ngũ (nghĩa là, hành động và tình trạng độc lập với nhau)

H_a: Café gây ra hiện tượng mất ngũ (nghĩa là, hành động và tình trạng không độc lập với nhau)

Bước 2: Tính giá trị χ^2 theo công thức

$$\chi^2 = \frac{N\left[(ad-bc)-\frac{N}{2}\right]^2}{R_1R_2C_1C_2}$$

$$\chi^2 = \frac{168\left[(30)(75)-(15)(48)-\frac{168}{2}\right]^2}{(78)(90)(45)(123)} = 9{,}04$$

Bước 3: So sánh giá trị χ^2 tính với giá trị χ^2 bảng ở độ tự do (r - 1)(c - 1), và loại giả thuyết độc lập nếu giá trị χ^2 tính lớn hơn giá trị χ^2 bảng ở mức ý nghĩa đã định.

Trong ví dụ, giá trị χ^2 bảng (Phụ lục V) với độ tự do bằng 1 là 3,84 ở mức ý nghĩa 5% và 6,63 ở mức ý nghĩa 1%. Vì giá trị χ^2 tính lớn hơn giá trị χ^2 bảng ở mức ý nghĩa 1%, nên giả thuyết tính độc lập giữa tình trạng và hành động bị bác bỏ. Từ đó có thể kết luận café gây ra hiện tượng mất ngũ.

b. Trường hợp bảng liên hợp 2 x 3

Giá trị χ^2 được tính như sau.

$$\chi^2 = \frac{N^2}{R_1R_2}\sum_{i,j=1}^{r,c}\frac{o_{ij}^2}{C_j} - \frac{NR_1}{R_2}$$

với N là tổng số quan sát; R_1 và R_2 lần lượt là tổng hàng thứ nhất và thứ hai; o_{ij} là giá trị quan sát ở hàng thứ i và cột thứ j; và C_j là tổng cột thứ j.

Ví dụ: Một nhà côn trùng học muốn biết tỷ lệ sâu non mới nở giữa các thời kỳ quan sát có khác nhau hay không. Ông đã theo dõi các lứa sâu trong từng thời kỳ và đếm số sâu non mới nở bắt được như sau:

Thời kỳ theo dõi (Tháng)	1	2	3	4	5	Tổng hàng
Số sâu non mới nở bắt được	62	28	70	75	15	250
Số sâu non mới nở không bắt được	426	364	210	440	170	1610
Tổng số sâu bắt được	488	392	280	515	185	1860

Các bước thực hiện kiểm định:

Bước 1: Đặt các giả thuyết:

H_0: Tỷ lệ sâu non mới nở giữa các thời kỳ quan sát như nhau

H_a: Tỷ lệ sâu non mới nở giữa các thời kỳ quan sát khác nhau

Bước 2: Tính giá trị χ^2 theo công thức

$$\chi^2 = \frac{N^2}{R_1R_2}\sum_{i,j=1}^{r,c}\frac{o_{ij}^2}{C_j} - \frac{NR_1}{R_2}$$

$$\chi^2 = \frac{(1860)^2}{(250)(1610)}\left(\frac{62^2}{488}+\frac{28^2}{392}+\frac{70^2}{280}+\frac{75^2}{515}+\frac{15^2}{185}\right) - \frac{(1860)(250)}{1610}$$

$$= 50{,}83$$

Bước 3: So sánh giá trị χ^2 tính với giá trị χ^2 bảng, ở (r - 1)(c - 1) độ tự do. Trong ví dụ, giá trị χ^2 bảng ở độ tự do (2 – 1)(5 – 1) = 4 là 9,49 ở mức ý nghĩa 5% và 13,6 ở mức ý nghĩa 1%. Vì giá trị χ^2 tính được là 50,83 lớn hơn rất nhiều so với giá trị χ^2 bảng ở mức ý nghĩa 1%, nên có thể kết luận tỷ lệ sâu non mới nở giữa các thời kỳ quan sát khác nhau.

c. Trường hợp bảng liên hợp r x c

Giá trị χ^2 được tính theo công thức như sau:

$$\chi^2 = N\sum_{i,j=1}^{r,c}\frac{(o_{ij})^2}{R_iC_j} - 1$$

với N là tổng số quan sát; o_{ij} là giá trị quan sát ở hàng thứ i và cột thứ j; R_i là tổng hàng thứ i và C_j là tổng cột thứ j.

Ví dụ: Một nhà kinh tế nông nghiệp nghiên cứu các yếu tố ảnh hưởng đến việc chấp nhận các giống lúa mới, nhập nội, năng suất cao. Ông muốn biết việc chấp nhận này có bị ảnh hưởng bởi tình trạng chiếm hữu của nông dân hay không?

Ông đặt tiêu chuẩn phân nhóm thứ nhất là ba nhóm tình trạng chiếm hữu khác nhau: chủ điền, tá điền thuê đất theo vụ và tá điền thuê đất thường xuyên. Tiêu chuẩn phân nhóm thứ hai là hai nhóm tình trạng: chấp nhận và không chấp nhận giống lúa mới. Kết quả số liệu tạo thành bảng ngẫu nhiên 3 x 2. Đối với bộ số liệu như thế mỗi nông dân được lấy mẫu có thể xếp vào một trong sáu nhóm sau:

(1) Chủ điền - chấp nhận
(2) Chủ điền - không chấp nhận
(3) Tá điền thuê đất theo vụ - chấp nhận
(4) Tá điền thuê đất theo vụ - không chấp nhận
(5) Tá điền thuê đất thường xuyên - chấp nhận
(6) Tá điền thuê đất thường xuyên - không chấp nhận.

Sử dụng số liệu dưới đây để kiểm định giả thuyết tính độc lập giữa hai tiêu chuẩn phân nhóm trong bảng ngẫu nhiên 3 x 2.

Tình trạng chấp nhận giống lúa mới

Tình trạng chiếm hữu	Số nông dân		Tổng hàng (R)
	Chấp nhận	Không chấp nhận	
Chủ điền	102	26	128
Tá điền thuê đất theo vụ	42	10	52
Tá điền thuê đất thường xuyên	4	3	7
Tổng cột (C)	148	39	187 (N)

Các bước thực hiện kiểm định:

Bước 1: Đặt các giả thuyết:

H_0: Tình trạng chấp nhận giống lúa mới của nông dân độc lập với tình trạng chiếm hữu

H_a: Tình trạng chấp nhận giống lúa mới của nông dân không độc lập với tình trạng chiếm hữu

Bước 2: Tính giá trị χ^2 :

$$\chi^2 = N\sum_{i,j=1}^{r,c}\frac{(o_{ij})^2}{R_iC_j} - 1$$

$$\chi^2 = 187\left[\frac{(102)^2}{128x148} + \frac{(26)^2}{128x39} + \cdots + \frac{3^2}{7x39}\right] - 1 = 2{,}01$$

Bước 3: So sánh giá trị χ^2 tính với giá trị χ^2 bảng, ở (r - 1)(c - 1) độ tự do. Trong ví dụ, giá trị χ^2 bảng (Phụ lục V) với (r - 1)(c - 1) = (2)(1) = 2 độ tự do là 5,99 ở mức ý nghĩa 5% và 9,21 ở mức ý nghĩa 1%. Vì giá trị χ^2 tính nhỏ hơn giá trị χ^2 bảng ở mức ý nghĩa 5%, nên giả thuyết về tính độc lập giữa việc chấp nhận giống lúa mới và tình trạng chiếm hữu của nông dân không thể bác bỏ được.

Trong phương pháp kiểm định tính độc lập giữa hai tiêu chuẩn phân nhóm, không có sự đánh giá về bản chất của phân bố tần số cho mỗi tiêu chuẩn. Nếu muốn làm điều này nên áp dụng kiểm định cho giả thuyết có tỷ lệ cố định cho từng tiêu chuẩn tùy thuộc vào kết quả của kiểm định tính độc lập. Nghĩa là, nếu các tiêu chuẩn phân nhóm khác nhau được đánh giá độc lập, lúc đó kiểm định cho giả thuyết có tỷ lệ cố định: hoặc được thực hiện trên biến hàng hoặc trên biến cột hoặc trên biến tổng (tổng hàng cho biến hàng và tổng cột cho biến cột). Tuy nhiên, nếu hai tiêu chuẩn phân nhóm không độc lập, lúc đó kiểm định giả thuyết có tỷ lệ cố định nên được áp dụng cho biến hàng ở mỗi cột, và cho biến cột ở mỗi hàng.

Trong ví dụ, vì hai tiêu chuẩn phân nhóm được đánh giá độc lập, nên kiểm định chi bình phương cho giả thuyết có tỷ lệ cố định nên được thực hiện trên biến tổng. Giả sử nghiên cứu viên cần quan tâm đến việc kiểm định tỷ lệ 2 : 1 cho tiêu chuẩn chấp nhận (nghĩa là biến cột), kiểm định chi bình phương cho giả thuyết có tỷ lệ cố định (mục I.1.a) được áp dụng cho tổng cột sẽ được tính như sau:

$$\chi^2 = \frac{(|148 - 124{,}7| - 0{,}5)^2}{124{,}7} + \frac{(|39 - 62{,}3| - 0{,}5)^2}{62{,}3} = 12{,}51$$

Vì giá trị χ^2 tính lớn hơn giá trị χ^2 bảng tương ứng ở mức ý nghĩa 1%, nên giả thuyết H_0 (nghĩa là, tần số của người chấp nhận giống lúa mới gấp đôi tần số của người không chấp nhận) bị bác bỏ ở mức ý nghĩa 1%.

3. Kiểm định tính đồng nhất của tỷ lệ *(Test for Homogeneity of ratio)*

Đối với các thí nghiệm được lặp lại nhiều lần. Ví dụ, nghiên cứu về tính ưa thích của rầy xanh ở lúa có thể được lặp lại qua nhiều thời điểm khác nhau hoặc có thể lặp lại trong nhiều lồng cùng một thời gian.

Khi có được số liệu của nhiều thí nghiệm, cần xác định cách nào để gom thông tin của tất cả thí nghiệm lại. Chẳng hạn, để nghiên cứu tính ưa thích của rầy, nhà nghiên cứu thực hiện nhiều thí nghiệm riêng biệt, các câu hỏi có thể được đặt ra là: (1) Có cơ sở để cộng dễ dàng số côn trùng trên cây bệnh và trên cây mạnh của tất cả thí nghiệm và áp dụng chỉ một lần phương pháp kiểm định chi bình phương cho tỷ lệ được giả thuyết là 1:1 dựa trên các giá trị tổng? (2) hoặc có nên áp dụng kiểm định chi bình phương cho kết quả của từng thí nghiệm riêng biệt? (3) Nếu như thế, làm thế nào tất cả kết quả kiểm định sẽ được giải thích và được phối hợp lại? Kiểm định chi bình phương về tính đồng nhất của tỷ lệ sẽ thích hợp để trả lời các câu hỏi này.

Để giải thích phương pháp áp dụng kiểm định tính đồng nhất của một tỷ lệ, chúng ta dùng số liệu của bốn thí nghiệm nghiên cứu về tính ưa thích của rầy xanh (Bảng 10.1).

Các bước thực hiện kiểm định:

Bước 1: Đặt các giả thuyết:

H_0: Các thí nghiệm đồng nhất một tỷ lệ.

H_a: Các thí nghiệm không đồng nhất một tỷ lệ.

Bước 2: Áp dụng kiểm định χ^2 cho giả thuyết có tỷ lệ cố định cho từng thí nghiệm riêng biệt (xem mục I.1.a). Trong ví dụ, chú ý giá trị χ^2 của thí nghiệm đầu tiên được tính trong mục I.1.a. Giá trị χ^2 của các thí nghiệm thứ hai, thứ ba và thứ tư được tính tương tự như thí nghiệm đầu tiên. Kết quả của bốn thí nghiệm được trình bày ở cột cuối cùng của Bảng 10.1.

Bảng 10.1: Áp dụng kiểm định chi bình phương cho tính đồng nhất của tỷ lệ 1:1 về số lượng rầy xanh được tìm thấy trên cây lúa bệnh và cây mạnh trong bốn thí nghiệm.

Số thí Nghiệm	Tổng số côn trùng	Giá trị quan sát mạnh	Giá trị quan sát bệnh	Giá trị mong đợi mạnh	Giá trị mong đợi bệnh	Giá trị χ^2
1	239	67	172	119,5	119,5	45,26
2	183	74	109	91,5	91,5	6,32
3	171	54	117	85,5	85,5	22,48
4	301	97	204	150,5	150,5	37,33
Tổng cộng	**894**	**292**	**602**	**447,0**	**447,0**	**106,80 = χ^2_T**
Tổng χ^2						***111,39 = χ^2_S***

Bước 3: Tính tổng của tất cả giá trị χ^2 được tính ở bước 1 như sau:

$$\chi^2_S = \sum_{i=1}^{S} \chi^2_i$$

với χ^2_i là giá trị χ^2 tính được của thí nghiệm thứ i và s là tổng số thí nghiệm.

Trong ví dụ, tổng của 4 trị số χ^2 là:

$\chi^2_S = \chi^2_4 = 45,26 + 6,32 + 22,48 + 37,33 = 111,39$

Bước 4: Tính tổng các giá trị quan sát và tổng các giá trị mong đợi của tất cả thí nghiệm, áp dụng cùng phương pháp kiểm định χ^2 được dùng ở bước 1 đối với các số tổng cộng này. Trong ví dụ, các số tổng cộng được trình bày ở hàng thứ năm của Bảng 10.1 và tính giá trị χ^2 dựa trên các số tổng cộng này như sau:

$$\chi^2_T = \frac{(|292-447|-0,5)^2}{447} + \frac{(|602-447|-0,5)^2}{447} = 106,80$$

Bước 5: Tính sự sai biệt giữa giá trị χ^2_S và χ^2_T

$\chi^2_d = \chi^2_S - \chi^2_T$

$\chi^2_d = 111,39 - 106,80 = 4,59$

Bước 6: So sánh giá trị χ^2_d với giá trị χ^2 bảng, với (s - 1) độ tự do ở mức ý nghĩa α và chọn cách làm sau đây tùy thuộc vào kết quả:

* Nếu giá trị χ^2_d tính được nhỏ hơn hoặc bằng giá trị χ^2 bảng, chứng tỏ các thí nghiệm đồng nhất một tỷ lệ. Như vậy, số liệu của tổng s thí nghiệm có thể được gom chung lại và sử dụng giá trị χ^2_S làm tiêu chuẩn kiểm định để kiểm định giả thuyết của tỷ lệ đã được giả định. Nghĩa là, chỉ có giá trị χ^2_S duy nhất cần để so sánh với giá trị χ^2 bảng với s độ tự do ở mức ý nghĩa α.

* Nếu giá trị χ^2_d tính được lớn hơn giá trị χ^2 bảng, chứng tỏ số liệu của s thí nghiệm không đồng nhất (nghĩa là, s bộ số liệu không chia đều một tỷ lệ chung); do đó, số liệu của s thí nghiệm không thể gom chung lại. Trong trường hợp như thế, nên xem xét từng giá trị χ^2 và xác định những nguyên nhân có thể có đối với sự khác biệt về tỷ lệ giữa các thí nghiệm.

Theo ví dụ, giá trị χ^2 bảng, với (s - 1) = (4 - 1) = 3 độ tự do là 7,81 ở mức ý nghĩa 5% và 11,34 ở mức ý nghĩa 1%. Vì giá trị χ^2_d nhỏ hơn giá trị χ^2 bảng ở mức ý nghĩa 5%. Do đó, số liệu của bốn thí nghiệm có thể được gom chung lại và so sánh giá trị χ^2_4 với giá trị χ^2 bảng ở độ tự do bằng 4.

Vì giá trị χ^2_4 là 111,39 lớn hơn giá trị χ^2 bảng ở mức ý nghĩa 1% là 13,28, nên giả thuyết của tỷ lệ 1 : 1 bị loại. Kết quả này chứng tỏ rằng rầy xanh hiện diện trên cây lúa

bệnh nhiều hơn cây mạnh và giả thuyết tính không ưa thích bị bác bỏ. Thật vậy, số liệu dường như cho thấy tỷ lệ khoảng 2 : 1 nghiêng về các cây bị bệnh.

II. KIỂM ĐỊNH CHO TÍNH ĐỒNG NHẤT CỦA PHƯƠNG SAI

Các phương pháp kiểm định thống kê đã được thảo luận đều dựa trên tính giả thuyết phổ biến nhất là tính bằng nhau của các trung bình nghiệm thức. Trong phần này, chúng ta sẽ thảo luận phương pháp thống kê để kiểm định tính bằng nhau (hoặc tính đồng nhất) của nhiều phương sai. Đây là kiểm định chi bình phương về tính đồng nhất của phương sai, còn gọi là kiểm định Bartlett (Bartlett's test). Trong nghiên cứu nông nghiệp, kiểm định này thường được sử dụng để:

* Kiểm tra tính đồng nhất của các phương sai như là một yêu cầu để phân tích phương sai có giá trị.

* Kiểm tra tính đồng nhất của các phương sai sai số trong việc phối hợp số liệu của một loạt các thí nghiệm.

* Kiểm tra tính đồng nhất của các phương sai trong nghiên cứu di truyền ở đó các vật liệu thí nghiệm bao gồm các kiểu gen của các thế hệ con cái khác nhau.

* Kiểm tra tính đồng nhất của các phương sai lấy giữa các mẫu được rút ra từ hai hay nhiều tổng thể.

Kiểm định chi bình phương cho tính đồng nhất của nhiều phương sai được áp dụng bất cứ lúc nào có nhiều hơn hai phương sai được kiểm định (kiểm định Bartlett). Khi chỉ có hai phương sai, nên sử dụng kiểm định F (kiểm định Hartley) với giá trị F được tính bằng tỉ số của hai phương sai – phương sai có giá trị lớn ở tử số và phương sai có giá trị nhỏ ở mẫu số.

Công thức tính χ^2 trong kiểm định Bartlett như sau:

$$\chi^2 = \frac{1}{C}(n-1)\left(k\log_e s_p^2 - \sum_{i=1}^{k}\log_e s_i^2\right)$$

Công thức trên có thể chuyển sang dạng logarith thập phân:

$$\chi^2 = \frac{1}{C}.2{,}3026.(n-1)\left(k\log_{10} s_p^2 - \sum_{i=1}^{k}\log_{10} s_i^2\right)$$

với k = số phương sai (số thí nghiệm)

n = số lần lặp lại (số quan sát) ở mỗi thí nghiệm

2,3026 = hằng số đúng cho mọi trường hợp sử dụng logarith thập phân ($\log_e 10 = 2{,}3026$)

$$C = 1 + \frac{k+1}{3k(n-1)}$$

s_p^2 = Phương sai gộp (phương sai trung bình của k thí nghiệm)

$$s_p^2 = \frac{\sum_{i=1}^{k} s_i^2}{k}$$

Chúng ta sử dụng hai trường hợp để giải thích phương pháp áp dụng kiểm định chi bình phương cho tính đồng nhất của các phương sai. Một trường hợp cho tất cả phương sai được ước lượng có độ tự do bằng nhau và một trường hợp khác cho các ước lượng phương sai có độ tự do không bằng nhau.

1. Độ tự do bằng nhau

Sử dụng số liệu của thí nghiệm với 11 giống lúa được trắc nghiệm ở ba chế độ nhiệt khác nhau trong phòng tăng trưởng để giải thích kiểm định cho tính đồng nhất của các phương sai có độ tự do bằng nhau. Ở mỗi nhiệt độ, các giống được trồng theo kiểu bố trí khối hoàn toàn ngẫu nhiên với ba lần lặp lại. Để phối hợp các số liệu của ba thí nghiệm (mỗi thí nghiệm tương ứng một nhiệt độ), cần kiểm tra tính đồng nhất của các phương sai sai số từ ba phân tích phương sai RCBD riêng biệt. Số liệu được thu thập là chiều cao cây (cm) và ba trung bình bình phương sai số, mỗi trung bình bình phương có 20 độ tự do:

$s_1^2 = 11,459848$

$s_2^2 = 17,696970$

$s_3^2 = 10,106818$

Để kiểm định tính đồng nhất của k phương sai (k = 3) có độ tự do bằng nhau, các bước được thực hiện như sau:

Bước 1: Tính log s^2 (log là logarith thập phân), các giá trị tổng của k giá trị s^2 và log s^2.

	$\log s^2$	s^2
1	1,059179	11,459848
2	1,247899	17,696970
3	1,004614	10,106818
Tổng cộng	3.311692	39,263636

Bước 2: Ước lượng phương sai gộp như sau:

$$s_p^2 = \frac{\sum_{i=1}^{k} s_i^2}{k} = \frac{39,263636}{3} = 13,087879$$

Bước 3: Gọi (n - 1) là độ tự do của mỗi s_i^2,tính giá trị χ^2:

$$\chi^2 = \frac{(2,3026).(n-1)\left(k \log s_p^2 - \sum_{i=1}^{k} \log s_i^2\right)}{1 + [(k+1)/3k(n-1)]}$$

Trong ví dụ, giá trị χ^2 được tính như sau:

$$\chi^2 = \frac{(2,3026)(20)[(3)(\log 13,087879) - 3,311692]}{1 + [(3+1)/(3)(3)(20)} = 1,75$$

Bước 4: So sánh giá trị χ^2 tính với giá trị χ^2 bảng ở (k - 1) độ tự do; và bác bỏ giả thuyết phương sai đồng nhất nếu giá trị χ^2 tính vượt quá giá trị χ^2 bảng ở mức ý nghĩa α.

Trong ví dụ, vì giá trị χ^2 tính nhỏ hơn giá trị χ^2 bảng với df = (k - 1) = (3 - 1) ở mức ý nghĩa 5% là 5,99, nên giả thuyết ba phương sai sai số đồng nhất không bị bác bỏ.

2. Độ tự do không bằng nhau

Phương pháp kiểm định tính đồng nhất của các phương sai có độ tự do không bằng nhau được giải thích bằng cách sử dụng số liệu về chiều dài vết bệnh được thu thập từ các lá lúa được tiêm chủng với bốn dòng phân lập khác nhau. Đối với mỗi dòng phân lập, đo 17 - 20 vết bệnh được chọn ngẫu nhiên. Nhà nghiên cứu muốn xác định có sự khác biệt về chiều dài của vết bệnh giữa các dòng phân lập khác nhau không? Trước khi áp dụng kiểm định để so sánh trung bình, đầu tiên phải chứng minh được tính đồng

nhất của các phương sai lấy mẫu giữa các dòng phân lập. Số liệu của bốn phương sai lấy mẫu và độ tự do tương ứng của chúng được trình bày trong Bảng 10.2.

Bảng 10.2: Các tham số dùng để kiểm định tính đồng nhất của các phương sai có độ tự do không bằng nhau

Dòng phân lập	Phương sai lấy mẫu (s^2)	Độ tự do (ν)	(s^2)(ν)	$\log s^2$	($\log s^2$)(ν)	$\frac{1}{\nu_i}$
1	6,73920	19	128,0448	0,828608	15,743552	0,0526
2	1,93496	16	30,9594	0,286672	4,586752	0,0625
3	1,15500	17	19,6350	0,062582	1,063894	0,0588
4	10,58450	19	201,1055	1,024670	19,468730	0,0526
Tổng cộng		71	379,7447		40,862928	0,2265
Gộp	5,34852 (s_p^2)			0,728234	51,704614	

Các bước kiểm định tính đồng nhất của bốn phương sai lấy mẫu có độ tự do không bằng nhau được thực hiện như sau:

Bước 1: gọi ν_i là độ tự do của phương sai thứ i (i = 1,...,k), tính các tham số:

$(\nu_i)(s_i^2)$, $(\nu_i)(\log s_i^2)$, $\frac{1}{\nu_i}$ và tổng của k giá trị cho mỗi tham số.

Trong ví dụ, các tham số này được tính và trình bày trong Bảng 10.2

Bước 2: Tính số ước lượng của phương sai gộp như sau:

$$s_p^2 = \frac{\sum_{i=1}^{k}(s_i^2)(\nu_i)}{\sum_{i=1}^{k}\nu_i} = \frac{379,7447}{71} = 5,24852$$

Bước 4: Tính giá trị χ^2 :

$$\chi^2 = \frac{2,3026[(A)(\log s_p^2) - C]}{1 + \frac{1}{3(k-1)}\left(D - \frac{1}{A}\right)}$$

với $A = \sum_{i=1}^{k}(\nu_i)$, $C = \sum_{i=1}^{k}(\log s_i^2)(\nu_i)$ và $D = \sum_{i=1}^{k}\frac{1}{\nu_i}$

Trong ví dụ, giá trị χ^2 được tính như sau:

$$\chi^2 = \frac{2,3026[(71)(0,728234) - 40,862928]}{1 + \frac{1}{3(4-1)}\left(0,2265 - \frac{1}{71}\right)} = \frac{24,964}{1,024} = 24,38$$

Bước 5: So sánh giá trị χ^2 tính với giá trị χ^2 bảng ở (k - 1) độ tự do, và bác bỏ giả thuyết phương sai đồng nhất nếu giá trị χ^2 tính vượt quá giá trị χ^2 bảng ở mức ý nghĩa α.

Trong ví dụ, giá trị χ^2 bảng với df = 3 - 1 = 2 là 7,81 ở mức ý nghĩa 5% và 11,34 ở mức ý nghĩa 1%. Vì giá trị χ^2 tính lớn hơn giá trị χ^2 bảng ở mức ý nghĩa 1%, nên giả thuyết phương sai đồng nhất bị bác bỏ; nghĩa là, kiểm định cho thấy có sự khác nhau đáng kể giữa bốn phương sai mẫu.

III. KIỂM ĐỊNH CHO ĐỘ LỆCH CHUẨN CỦA TỔNG THỂ

(Test for a population standard deviation)

1. Kiểm định cho độ lệch chuẩn của tổng thể

Để kiểm định giả thuyết cho độ lệch chuẩn của tổng thể với giả thuyết không,

H_0: $\sigma = \sigma_G$. (Tổng thể có phân bố chuẩn).

H_a: $\sigma = \sigma_G$. (Tổng thể khác phân bố chuẩn).

Ví dụ: Một công ty chế tạo đồ gia dụng nhà sản xuất chai có đường kính 10mm. Công ty biết rằng đường kính của chai được sản xuất có sự biến động ít nhiều quanh giá trị 10 mm và cũng thay đổi từ chai này sang chai khác. Nhưng dù có chấp nhận sự biến động của đường kính chai, công ty cũng không thể chịu đựng sự biến động quá lớn Vì như thế sẽ có quá nhiều chai được sản xuất không thể sử dụng được.

Để biết chắc độ lệch chuẩn (σ) của đường kính chai không quá lớn. Do không thể biết được giá trị σ một cách chính xác, công ty áp dụng các phương pháp suy luận như sau:

Giả sử độ lệch chuẩn có thể chấp nhận cho đường kính chai nhỏ hơn 0,09 mm ($\sigma < 0{,}09$ mm). Để kết luận đường kính chai được sản xuất có biến động quá nhiều hay ít bằng cách thực hiện kiểm định giả thuyết.

Bước 1: Đặt các giả thuyết:

H_0: $\sigma = 0{,}09$ mm (biến động quá nhiều)

H_a: $\sigma < 0{,}09$ mm (không biến động quá nhiều)

Như vậy, nếu loại giả thuyết không; lúc đó, công ty tin tưởng biến động của đường kính chai có thể chấp nhận được.

Để thực hiện kiểm định giả thuyết, đầu tiên cần lấy mẫu ngẫu nhiên các chai, sau đó tính độ lệch chuẩn (s) của đường kính các chai được lấy mẫu. Nếu s < 0,09 quá nhiều, giả thuyết không sẽ bị loại và chấp nhận giả thuyết chọn lựa; trái lại, không loại giả thuyết không.

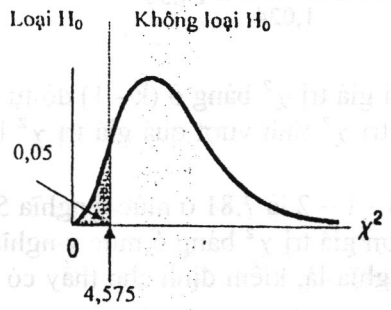
Công ty lấy một mẫu ngẫu nhiên gồm 12 chai và đo đường kính của từng chai. Kết quả như sau:

10,00	9.92	10,00	10,00
9,98	10.02	10,12	9,97
10,03	9,98	10,06	10,05

Thực hiện kiểm định đuôi trái χ^2_{1-2}, với df = n - 1 ở mức ý nghĩa 5%.

Bước 2: Tìm giá trị tới hạn.

Vì $\alpha = 0{,}05$, n = 12 nên df = 12 - 1 = 11. Tìm trong Phụ lục V, giá trị chuẩn $\chi^2_{1-\alpha} = \chi^2_{1-0,05} = \chi^2_{0,95} = 4{,}575$ (Hình 10.1)



Hình 10.1: Tiêu chuẩn để quyết định có nên loại giả thuyết không ?

Tính độ lệch chuẩn của mẫu, s:

$$s = \sqrt{\frac{\sum X^2 - (\sum X)^2}{n-1}} = \sqrt{\frac{1202,6299 - (120,13)^2}{11}} = 0,051 \text{ mm}$$

$s^2 = 0.0026$

Bước 3: Tính số thống kê kiểm định

$$\chi^2 = \frac{n-1}{\sigma_G^2} s^2$$

với n = 12 và σ_G = 0,09, giá trị của số thống kê kiểm định là

$$\chi^2 = \frac{n-1}{\sigma_G^2} s^2 = \frac{12-1}{(0,09)^2}.0,0026 = 3,531$$

Nếu giá trị χ^2 nằm trong vùng bác bỏ, lúc đó bác bỏ giả thuyết; ngược lại, không bác bỏ giả thuyết H_0.

Do giá trị của số thống kê kiểm định $\chi^2 = 3,531$ nằm trong vùng bác bỏ, nên chúng ta bác bỏ giả thuyết H_0. Như vậy, số liệu cung cấp đủ dẫn chứng để kết luận độ lệch chuẩn, σ, của đường kính của tất cả các chai 10 mm được hãng sản suất thì nhỏ hơn 0,09 mm. Chúng ta có thể kết luận biến động của đường kính chai không quá lớn.

2. Khoảng tin cậy cho độ lệch chuẩn của tổng thể

Để tìm khoảng tin cậy cho độ lệch chuẩn của tập hợp, σ. Chúng ta giả định tập hợp có phân bố chuẩn.

Cũng với ví dụ trên để xác định khoảng tin cậy 95% cho độ lệch chuẩn (σ) của đường kính tất cả các chai 10 mm được công ty sản suất. Chúng ta thực hiện như sau:

Đối với khoảng tin cậy 95%, mức tin cậy là 0,95 = 1 - 0,05 (α = 0,05). Vì n = 12, df = 12 - 1 = 11. Tìm trong phụ lục V, chúng ta có:

$$\chi^2_{1-\alpha/2} = \chi^2_{1-0,05/2} = \chi^2_{0,975} = 3,816$$

và

$$\chi^2_{\alpha/2} = \chi^2_{\alpha/2} = \chi^2_{0,025} = 21,920$$

Vậy, khoảng tin cậy cho σ được tính từ công thức sau:

$$\sqrt{\frac{n-1}{\chi^2_{\alpha/2}}}.s < \sigma < \sqrt{\frac{n-1}{\chi^2_{1-\alpha/2}}}.s$$

với n = 12, $\chi^2_{\alpha/2} = 21,920$, $\chi^2_{1-\alpha/2} = 3,816$ và s = 0,051 mm. Khoảng tin cậy 95% cho σ là

$$\sqrt{\frac{12-1}{21,920}}.0,051 < \sigma < \sqrt{\frac{12-1}{3,816}}.0,051$$

$$0,036 < \sigma < 0,087$$

Như vậy, chúng ta có thể tin cậy 95% độ lệch chuẩn của đường kính của tất cả các chai 10 mm được công ty sản xuất biến động giữa 0,036 và 0,087 mm.

IV. KIỂM ĐỊNH CHO TÍNH PHÙ HỢP *(Test for goodness-of-fit)*

Kiểm định tính phù hợp xác định xem bộ số liệu quan sát có phù hợp với phân bố xác suất đã định. Ví dụ, số liệu năng suất lúa hoặc hàm lượng protein có thể nghi ngờ tuân theo phân bố chuẩn. Phân bố của cỏ dại trong ruộng lúa hoặc của côn trùng bắt được trong bẫy qua thời gian có thể nghi ngờ có phân bố Poisson.

Trong nghiên cứu để xác định kiểu di truyền của hàm lượng protein của lúa, người ta đo phần trăm protein trong hạt các cây F_3 của tổ hợp lai giữa hai giống lúa nâu khác nhau. Nhìn dạng phân bố của số liệu ở hình 10.2, chúng ta thấy dường như lệch thật sự khỏi phân bố chuẩn. Do đó, nhà nghiên cứu muốn kiểm tra xem sự quan sát bằng mắt có thể được chứng minh bằng kiểm định thống kê không? Với mục đích này, kiểm định chi bình phương cho tính phù hợp của số liệu quan sát đối với phân bố chuẩn được giả định là thích hợp nhất.

Các bước thực hiện bao gồm:

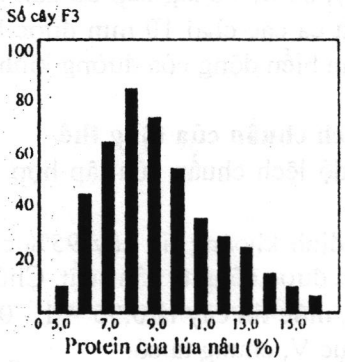
Bước 1: Đặt các giả thuyết không và chọn lựa:

H_0 : Phân bố của hàm lượng protein ở các giống lúa nâu có dạng phân bố chuẩn

H_a: Phân bố của hàm lượng protein ở các giống lúa nâu khác dạng phân bố chuẩn

Bước 2: Gọi n là số quan sát trong bộ số liệu, lập bảng tần số theo cách sau:

(1) Xác định khoảng biến động của số liệu, đó là khoảng chênh lệch giữa giá trị quan sát lớn nhất và nhỏ nhất, sau đó chia khoảng biến động thành p nhóm theo hướng dẫn dưới đây:



Hình 10.2: Phân bố tần số của hàm lượng protein ở giống lúa nâu trên 450 cây F_3

* Số nhóm (p) nên ở trong khoảng 8 - 18 tùy theo số quan sát (n) và khoảng biến động. Số nhóm p được khuyến cáo cho các khoảng biến động khác nhau của n như sau:

Số quan sát (n)	Số nhóm (p)
20 - 100	8 - 12
101 - 500	10 - 15
501 - 1000	12 - 18

Lưu ý, tùy theo khoảng biến động của bộ số liệu nên chọn p ở một trong ba cách trên. Số liệu càng biến động nên chọn p càng lớn.

* Không nên có nhóm không có số quan sát và nên tránh các nhóm có ít hơn ba số quan sát, đặc biệt các nhóm gần giữa khoảng biến động. Khi gặp các nhóm có ít hơn ba số quan sát ở giữa khoảng biến động, nên giảm số nhóm. Khi các nhóm có ít hơn ba số quan sát xuất hiện gần một trong hai đầu của khoảng biến động, có thể phối hợp lại hai hoặc ba nhóm kế nhau.

* Để dễ tính toán, cần định rõ các giới hạn trên và dưới của nhóm để các điểm giữa của khoảng biến động trong nhóm (nghĩa là, các giá trị giữa của nhóm) càng gần với số nguyên càng tốt.

Giá trị giữa của nhóm được tính bằng cách lấy trung bình của các giới hạn trên và dưới của nhóm đó.

Trong ví dụ, với n = 450 và số liệu biến động từ 4,6 - 16,3, chúng ta có thể chia khoảng biến động thành p = 12 nhóm. Các giá trị giữa của 12 nhóm được ghi trong cột thứ ba của Bảng 10.3.

(2) Xác định tần số quan sát (o_i), đó là số biến quan sát được trong nhóm thứ i. Tổng tần số quan sát của tất cả các nhóm $\sum_{i=1}^{p} o_i$ nên bằng với tổng số biến quan sát (n).

$$\sum_{i=1}^{p} o_i = n = 450$$

Bước 3: Tính trung bình ($\overline{X}$) và phương sai (s^2) từ Bảng 10.3

$$\overline{X} = \frac{\sum_{i=1}^{p}(X_i)(o_i)}{n} = 4049/450 = 8,998$$

$$s^2 = \frac{\sum_{i=1}^{p}(X_i^2)(o_i) - \frac{\left[\sum_{i=1}^{p}(X_i)(o_i)\right]^2}{n}}{n-1} = \frac{38,757-(4049)^2/450}{450-1} = 5,178$$

với X_i là giá trị giữa của nhóm i, o_i là tần số quan sát của nhóm i và p là số nhóm.

Bảng 10.3: Phân bố tần số, tính trung bình và phương sai của số liệu protein lúa nâu của 450 cây F_3.

Nhóm			Tần số quan sát	Tính trung bình và phương sai		
Số	Giới hạn	Giá trị (Xi)	(o_i)	(X_i)(o_i)	X_i^2	(X_i^2)(o_i)
1	4,5 - 5,4	5,0	7	35	25	175
2	5,5 - 6,4	6,0	5	312	36	1872
3	6,5 - 7,4	7,0	70	490	49	3430
4	7,5 - 8,4	8,0	81	648	64	5184
5	8,5 - 9,4	9,0	74	666	81	5994
6	9,5 - 10,4	10,0	63	630	100	6300
7	10,5 - 11,4	11,0	39	429	121	4719
8	11,5 - 12,4	12,0	24	288	144	3456
9	12,5 - 13,4	13,0	21	273	169	3549
10	13,5 - 14,4	14,0	10	140	196	1960
11	14,5 - 15,4	15,0	6	90	225	1350
12	15,5 - 16,4	16,0	3	48	256	768
Tổng	-	-	**450**			**4049**

Bước 4: Tính tần số mong đợi của mỗi nhóm dựa trên phân bố xác suất lý thuyết:

(1) Đối với mỗi nhóm, tính hai giá trị chuẩn tắc (z): một cho giới hạn dưới (z_l) và một cho giới hạn trên (z_h):

$$z_l = \frac{L_l - \overline{X}}{s} \quad \text{và} \quad z_h = \frac{L_h - \overline{X}}{s}$$

với L_h và L_l lần lượt là các giới hạn trên và dưới thật của mỗi nhóm. Giới hạn dưới thật của một nhóm nào đó là trung bình của giới hạn dưới của chính nhóm đó và giới hạn trên của nhóm trước. Tương tự, giới hạn trên thật là trung bình của giới hạn trên của

chính nhóm đó và giới hạn dưới của nhóm kế tiếp. Giới hạn thấp thật của nhóm đầu tiên là - ∞ và giới hạn trên thật của nhóm sau cùng là + ∞.

Ví dụ: Giới hạn thấp thật (L_l) và giới hạn trên thật (L_h) của nhóm thứ hai trong ví dụ là:

$$L_l = \frac{5,4+5,5}{2} = 5,45 \quad \text{và} \quad L_h = \frac{6,4+6,5}{2} = 6,45$$

và giá trị z_l và z_h tương ứng được tính như sau:

$$z_l = \frac{5,45-8,998}{2,2755} = -1,559 \quad \text{và} \quad z_h = \frac{6,45-8,998}{2,2755} = -1,120$$

Các giá trị chuẩn tắc (z) của 12 nhóm được trình bày trong cột thứ hai và thứ ba của Bảng 10.4.

Bảng 10.4: Tính tần số mong đợi và kiểm định tính phù hợp với phân bố chuẩn từ số liệu ở Bảng 10.3.

Số nhóm	Trị số z		Xác suất (p_i)
	z_l	z_h	
1	-∞	-1,559	0,5000 - 0,4405 = 0,0595
2	-1,559	-1,120	0,4405 - 0,3686 = 0,0719
3	-1,120	-0,680	0,3686 - 0,2517 = 0,1169
4	-0,680	-0,241	0,2517 - 0,0952 = 0,1565
5	-0,241	0,199	0,0952 + 0,0789 = 0,1741
6	0,199	0,638	0,2383 - 0,0789 = 0,1594
7	0,638	1,078	0,3595 - 0,2382 = 0,1212
8	1,078	1,517	0,4353 - 0,3595 = 0,0758
9	1,517	1,956	0,4748 - 0,4353 = 0,0395
10	1,956	2,396	0,4917 - 0,4748 = 0,0169
11	2,396	2,835	0,4977 - 0,4917 = 0,0060
12	2,835	∞	0,5000 - 0,4977 = 0,0023

(2) Tính xác suất của mỗi khoảng nhóm, dựa trên phân bố xác suất lý thuyết, nghĩa là xác suất để z nằm giữa z_l và z_h, $p(z_l < z < z_h)$.

Trong ví dụ, do giả thuyết được đặt là phân bố chuẩn, nên xác suất của mỗi nhóm được xác định trong Phụ lục II bằng cách đọc diện tích dưới đường cong chuẩn tắc giữa z_l và z_h. Để thực hiện được điều nầy cần sử dụng các nguyên tắc sau:

* Độ chính xác của các giá trị z ghi ở Phụ lục II chỉ có 2 số lẻ. Vì vậy, khi các giá trị z có từ 3 số lẻ trở lên (như trong ví dụ), cần dùng phép nội suy để xác định diện tích mong muốn.

Ví dụ: Để xác định diện tích từ z = 0 đến z = 1,559, đầu tiên chúng ta tìm trong Phụ lục II diện tích từ 0 đến 1,55 là 0,4394 và diện tích từ 0 đến 1,56 là 0,4406. Qua phép nội suy, chúng ta tính diện tích từ 0 đến 1,559 như sau:

$$0,4394 + \frac{(0,4406-0,4394)(1559-1,55)}{(1,56-1,55)} = 0,4405$$

Nhắc lại:

* Diện tích giữa hai giá trị z bất kỳ có cùng dấu (z_l và z_h hoặc - z_l và - z_h) sẽ bằng hiệu số giữa các diện tích từ 0 đến z_l và từ 0 đến z_h. Ví dụ, để tính diện tích giữa hai giá trị z của nhóm 2 ở Bảng 10.4 (nghĩa là, z_l = -1,559 và z_h = -1,120), đầu tiên xác

định diện tích từ 0 đến 1,559 bằng 0,4405 và diện tích từ 0 đến 1,120 bằng 0,3686 (Phụ lục II), sau đó tính hiệu số của chúng:

$$P(-1,559 < z < -1,120) = P(1,559 < z < 1,120)$$
$$= 0,4405 - 0,3686 = 0,0719$$

* Diện tích giữa hai giá trị z bất kỳ trái dấu ($-z_l$ và z_h) sẽ bằng tổng diện tích từ 0 đến z_l và từ 0 đến z_h. Ví dụ, để tính diện tích giữa hai giá trị z của nhóm 5 (Bảng 10.4), nghĩa là $z_l = -0,241$ và $z_h = 0,199$. Bằng phép nội suy, đầu tiên xác định diện tích từ 0 đến 0,241 bằng 0,0952 và từ 0 đến 0,199 bằng 0,0789. Như vậy, diện tích giữa $z_l = -0,241$ và $z_h = 0,199$ là

$$p(-0,241 < z < 0,199) = 0,0952 + 0,0789 = 0,1741$$

* Diện tích từ $-\infty$ đến z, hoặc từ z đến $+\infty$ được tính bằng hiệu số giữa 0,5 và diện tích từ 0 đến z. Trong ví dụ, để tính diện tích từ $-\infty$ đến $z_h = -1,559$ của nhóm thứ nhất, đầu tiên tính diện tích từ 0 đến 1,559 bằng 0,4405, sau đó tính diện tích từ $-\infty$ đến -1,559 bằng 0,5 - 0.4405 = 0,0595.

Kết quả tính xác suất cho cả 12 nhóm được trình bày trong cột thứ tư của Bảng 10.4.

(3) Tính tần số mong đợi của nhóm i (e_i) bằng tích của xác suất ở nhóm i (P_i) và số quan sát (n):

$$e_i = (n)(p_i)$$

Bước 5: Tính giá trị χ^2:

$$\chi^2 = \sum \frac{(o_i - e_i)^2}{e_i}$$

Lập bảng để tính χ^2 như sau:

Nhóm	o_i	e_i	$o_i - e_i$	$(o_i - e_i)^2$	$(o_i - e_i)^2 / e_i$
1	7	26,775	-19,775	391,051	14,605
2	52	32,355	19,645	385.926	11,928
3	70	52,605	17,395	302,586	5,752
4	81	70,425	10,575	111,831	1,588
5	74	78,345	- 4,345	18,879	0,241
6	63	71,730	- 8,730	76,213	1,062
7	39	54,540	-15,540	241,492	4,428
8	24	34,110	-10,110	102,212	2,997
9	21	17,775	3,225	10,401	0,585
10	10	7,605	2,395	5,736	0,754
11	6	2,700	3,300	10,890	4,033
12	3	1,035	1,965	3,861	3,731
Tổng	**450 = n**	**450,000**			**51,704 = χ^2**

Bước 6: So sánh giá trị χ^2 tính với giá trị χ^2 bảng (Phụ lục V) với (p - 1) độ tự do, bác bỏ giả thuyết (H_0) nếu giá trị χ^2 tính cao hơn giá trị χ^2 bảng ở mức ý nghĩa α.

Trong ví dụ, giá trị tính χ^2 bảng với 11 độ tự do ở mức ý nghĩa 1% là 24,72. Vì giá trị χ^2 tính lớn hơn χ^2 giá trị bảng nên kiểm định cho thấy bộ số liệu quan sát không phù hợp với giả thuyết phân bố chuẩn.

CHƯƠNG 11

PHÂN TÍCH SỐ LIỆU TỪ MỘT LOẠT THÍ NGHIỆM VÀ BIẾN ĐỔI SỐ LIỆU

Năng suất cây trồng phụ thuộc vào kiểu gen, môi trường mà cây trồng được trồng trong đó, cũng như sự tương tác giữa kiểu gen và môi trường. Kiểu gen và một vài yếu tố môi trường như mức phân bón, mật độ cây trồng, kiểm soát dịch sâu hại,... nhà nghiên cứu có thể kiểm soát được. Tuy nhiên, các yếu tố môi trường khác như nắng, mưa, một vài đặc tính của đất nhìn chung cố định và khó thay đổi tại một địa điểm và mùa vụ nhất định. Vì thế, nếu thực hiện thí nghiệm một lần tại một địa điểm duy nhất, nhà nghiên cứu chỉ có thể thay đổi và đánh giá các yếu tố có thể kiểm soát, nhưng không thể đánh giá được các yếu tố môi trường vượt ra ngoài sự kiểm soát của nhà nghiên cứu.

Ảnh hưởng của các yếu tố môi trường không thể kiểm soát lên năng suất cây trồng cũng quan trọng bằng, nếu không nói là quan trọng hơn, ảnh hưởng của các yếu tố có thể kiểm soát và việc đánh giá cũng như định lượng các ảnh hưởng của chúng thì không kém cần thiết. Vì các yếu tố không thể kiểm soát được dự đoán thay đổi theo mùa vụ và địa điểm, và cũng vì những thay đổi này có thể đo được, nên các ảnh hưởng của chúng lên hiệu suất của nghiệm thức có thể được đánh giá. Trong nghiên cứu cây trồng, cách được sử dụng phổ biến nhất để đánh giá ảnh hưởng của các yếu tố môi trường không thể kiểm soát lên sự đáp ứng của cây trồng là lặp lại thí nghiệm ở nhiều địa điểm hoặc qua nhiều mùa vụ hoặc cả hai.

I. PHÂN TÍCH SỐ LIỆU TỪ MỘT LOẠT THÍ NGHIỆM

(Analysis of data from a series of experiments)

1. Phân tích qua nhiều mùa vụ *(Analysis over seasons)*

Đối với một cây trồng nhất định ở một địa điểm cụ thể, việc gieo trồng thường không được bố trí rải đều trong năm mà được tụ trung lại rõ rệt trong một vài mùa vụ được xác định thích hợp, và mùa vụ này được lặp đi lặp lại qua nhiều năm. Ví dụ, ở ĐBSCL đậu nành thường được canh tác trong vụ Xuân hè. Mặc dù nó có thể được trồng hai đến ba vụ trong năm, nhưng mùa vụ thường giữ nguyên rõ rệt về ngày gieo và các điểm đặc trưng của môi trường mong đợi. Do đó, mùa trồng trong năm được xem như *biến cố định*. Một kỹ thuật tốt có thể được nhận ra riêng biệt để phù hợp với mùa vụ cụ thể. Thật vậy, mục đích chính của phân tích phối hợp qua nhiều vụ là để khảo sát tương tác giữa mùa vụ và nghiệm thức, đồng thời để xác định sự cần thiết của việc khuyến cáo kỹ thuật khác nhau cho mỗi mùa trồng.

Để giải thích phương pháp phân tích số liệu của các thí nghiệm qua nhiều mùa vụ, chúng ta sử dụng thí nghiệm phân bón với năm mức độ đạm được thử nghiệm trên lúa qua hai vụ. Các thí nghiệm được bố trí theo thể thức RCBD với ba lần lặp lại. Số liệu năng suất hạt được ghi nhận trong Bảng 11.1 và kết quả phân tích phương sai của từng thí nghiệm được trình bày trong Bảng 11.2.

Gọi v là số vụ (v = 2)

t là số nghiệm thức (t = 5)

r là số lần lặp lại (r = 3)

Bảng 11.1 Năng suất hạt (t/ha) của lúa được trắc nghiệm với năm mức độ đạm qua hai mùa vụ. (Trích từ Bảng 8.1 của Gomez & Gomez. 1984. Trang 318)

Mức độ N (kg/ha)	Lặp lại I	II	III	Tổng (T)	Trung bình ($\bar{X}$)
		Mùa nắng			
0 (N_0)	4,891	2,577	4,541	12,009	4,003
60 (N_1)	6,009	6,625	5,672	18,306	6,102
90 (N_2)	6,712	6,693	6,799	20,204	6,735
120 (N_3)	6,458	6,675	6,636	19,769	6,590
150 (N_4)	5,683	6,868	5,692	18,243	6,081
Tổng				**88,531**	
		Mùa mưa			
0 (N_0)	4,999	3,503	5,356	13,858	4,619
60 (N_1)	6,351	6,316	6,582	19,249	6,416
90 (N_2)	6,071	5,969	5,893	17,933	5,978
120 (N_3)	4,818	4,024	5,813	14,655	4,885
150 (N_4)	3,436	4,047	3,740	11,223	3,741
Tổng				**76,918**	

Bảng 11.2: Phân tích phương sai (bố trí RCBD) của từng thí nghiệm phân bón trên lúa với năm nghiệm thức và ba lần lặp lại qua hai vụ.

Nguồn biến động	Độ tự do	Tổng bình phương	Trung bình bình phương	F (tính)
		Mùa nắng		
Lặp lại	2	0,018627	0,009314	
Đạm (N)	4	14,533384	3,633346	6,43*
Sai số	8	4,522162	0,565270	
		Mùa mưa		
Lặp lại	2	1,242944	0,621472	
Đạm (N)	4	13,869888	3,467472	10,91**
Sai số	8	2,541472	0,317684	

* = Khác biệt ở mức ý nghĩa 5%

** = Khác biệt ở mức ý nghĩa 1%

Các bước phân tích được thực hiện như sau:

Bước 1: Phát thảo bảng phân tích phương sai phối hợp qua nhiều mùa vụ, dựa trên bố trí thí nghiệm cơ bản được sử dụng. Trong ví dụ, phát thảo phân tích phương sai phối hợp qua nhiều mùa vụ dựa trên bố trí RCBD được trình bày ở Bảng 11.3.

Bước 2: Tính các tổng bình phương (SS) khác nhau trong phân tích phương sai phối hợp của bước 1:

• Tính SS lặp lại trong vụ *(Reps. within season SS)* bằng tổng của v SS lặp lại, và SS sai số gộp *(Pooled error SS)* bằng tổng của v SS sai số lấy từ kết quả phân tích phương sai từng vụ. Nghĩa là,

$$\text{SS lặp lại trong vụ} = \sum_{i}^{v}(\text{SSlặp lại})_i$$

$$\text{SS sai số gộp} = \sum_{i}^{v}(\text{SSsai số})_i$$

với $(\text{SSlặp lại})_i$ và $(\text{SSsai số})_i$ lần lượt là SS lặp lại và SS sai số trong phân tích phương sai của vụ thứ i.

Bảng 11.3: Phát thảo phân tích phương sai phối hợp qua nhiều vụ[a], dựa trên bố trí RCBD với t nghiệm thức và r lần lặp lại.

Nguồn biến động	Độ tự do	Trung bình bình phương	F tính
Mùa vụ	v – 1	MS_V	$\frac{MS_V}{MS_R}$
Lặp lại trong vụ	v(r – 1)	MS_R	
Nghiệm thức (T)	t -1	MS_T	$\frac{MS_T}{MS_E}$
V x T	(v – 1)(t – 1)	MS_{VxT}	$\frac{MS_{VxT}}{MS_E}$
Sai số gộp	v(r -1)(t -1)	MS_E	
Tổng cộng	vrt - 1		

[a]Mùa vụ được khảo sát như biến cố định.

Trong ví dụ, sử dụng các giá trị tổng bình phương ở Bảng 11.2, hai tổng bình phương đã nêu ở trên được tính như sau:

$$\text{SS lặp lại trong vụ} = \sum_{i}^{v}(\text{SSlặp lại})_i = (\text{SSlặp lại})_{\text{khô}} + (\text{SSlặp lại})_{\text{mưa}}$$

$$= 0{,}018627 + 1{,}242944 = 1{,}261571$$

$$\text{SS sai số gộp} = \sum_{i}^{v}(\text{SSsai số})_i = (\text{SSsai số})_{\text{khô}} + (\text{SSsai số})_{\text{mưa}}$$

$$= 4{,}522162 + 2{,}541472 = 7{,}063634$$

• Tính các giá trị SS khác: Mùa vụ (SS_V), nghiệm thức (SS_T) và tương tác giữa mùa vụ và nghiệm thức (SS_{VxT}), hoặc từ giá trị tổng hoặc từ giá trị trung bình.

- Dựa vào giá trị tổng: Tính các SS khác nhau theo phương pháp chuẩn như sau:

$$C.F. = \frac{\left(\sum_{i=1}^{v} G_i\right)^2}{vrt}$$

$$SS_V = \frac{\sum_{i=1}^{v} G_j^2}{rt} - C.F$$

$$SS_T = \frac{\sum_{j=1}^{t} T_j^2}{vr} - C.F.$$

$$SS_{VxT} = \sum_{i=1}^{s}\sum_{j=1}^{t}\frac{(VT)_{ij}^2}{r} - C.F - SS_V - SS_T$$

với G_i là tổng chung của vụ thứ i, T_j là tổng nghiệm thức thứ j qua tất cả v vụ, và (VT) là tổng nghiệm thức thứ j ở vụ thứ i.

- Dựa vào giá trị trung bình: Tính các SS khác nhau như sau:

$$C.F. = vrt.(\overline{G})^2$$

$$SS_V = rt.\left(\sum_{i=1}^{v} \overline{V}_i^2\right)^2 - C.F.$$

$$SS_T = vr.\left(\sum_{j=1}^{t} \overline{T}_j^2\right)^2 - C.F.$$

$$SS_{V\times T} = r.\left[\sum_{i=1}^{v}\sum_{j=1}^{t} (\overline{VT})_{ij}^2\right] - C.F. - SS_V - SS_T$$

với $\overline{G}$ là trung bình chung ($\overline{x}$), $\overline{V}_i$ là trung bình của vụ thứ i, $\overline{T}_j$ là trung bình của nghiệm thức thứ j, và $\overline{VT}_{ij}$ là trung bình của nghiệm thức thứ j ở vụ thứ i.

Chú ý, các giá trị trung bình dùng để tính SS nên có cùng số lẻ có nghĩa như được sử dụng trong các giá trị tổng.

Trong ví dụ, cách tính được thực hiện dựa trên giá trị tổng ở Bảng 11.1 như sau:

$$C.F. = \frac{(88,531 + 76,918)^2}{30} = 912,44572$$

$$SS_T = \frac{(88,531)^2 + (76,918)^2}{(3)(5)} - 912,44572 = 4,495392$$

$$SS_V = \frac{(25,867)^2 + (37,555)^2 + (38,137)^2 + (34,424)^2 + (29,466)^2}{6} - 912,44572$$

$$18,748849$$

$$SS_{T\times V} = \frac{(12,009)^2 + (18,306)^2 + \cdots + (11,223)^2}{3} - 912,44572 - 4,495392 - 18,748849$$

$$= 9,654423$$

Bước 3: Tính trung bình bình phương cho mỗi nguồn biến động bằng cách chia từng tổng bình phương cho độ tự do tương ứng. Trong ví dụ, kết quả tính được trình bày trong Bảng 11.4

Bảng 11.4: Phân tích phương sai phối hợp qua hai vụ được tính từ số liệu của Bảng 11.1 và 11.2.

Nguồn biến động	Độ tự do	Tổng bình phương	Trung bình bình phương	F (tính)
Mùa vụ (V)	1	4,495392	4,495392	•
Lặp lại trong vụ	4	1,261571	0,315393	
Đạm (N)	4	18,748849	4,687212	10,62**
V x N	4	9,654423	2,413606	5,47**
Sai số gộp	16	7,063634	0,441477	
Tổng cộng	29			

** Khác biệt ở mức ý nghĩa 1%

• Độ tự do của lặp lại trong vụ không đủ để kiểm định ý nghĩa có giá trị (df < 6).

Bước 4: Kiểm định tính đồng nhất của v MS sai số từ phân tích phương sai của từng thí nghiệm bằng cách áp dụng kiểm định F (nếu có hai phương sai sai số) hoặc kiểm định chi bình phương (nếu có nhiều hơn hai phương sai sai số). Thường các phương sai sai số có thể được xem như đồng nhất nếu MS sai số cao nhất không lớn hơn gấp ba lần MS sai số nhỏ nhất.

Trong ví dụ, do chỉ có hai MS sai số, nên áp dụng kiểm định F

$$F = \frac{\text{MSsai số lớn}}{\text{MSsai số nhỏ}} = \frac{0,565270}{0,317684} = 1,78$$

Vì giá trị F tính nhỏ hơn 3,44 tương ứng với giá trị F bảng có $df_1 = df_2 = 8$ ở mức ý nghĩa 5%, nên không thể loại giả thuyết H_0 (nghĩa là, giả thuyết các phương sai sai số đồng nhất qua các vụ).

Bước 5: Nếu các phương sai sai số không đồng nhất chuyển đến bước 7. Trái lại, nếu chúng đồng nhất tính các giá trị F để kiểm định các ảnh hưởng khác như được chỉ ở Bảng 11.3.

$$F_V = \frac{MS_V}{\text{MS lặp lại trong vụ}}$$

$$F_T = \frac{MS_T}{\text{MS sai số gộp}}$$

$$F_V = \frac{MS_{V \times T}}{\text{MSsai số gộp}}$$

Trong ví dụ, kiểm định tính đồng nhất ở bước 4 không ý nghĩa; nghĩa là, các phương sai sai số đồng nhất. Vì thế, các giá trị F được tính theo công thức trên như sau:

$$F_N = \frac{4,687212}{0,441477} = 10,62$$

$$F_{V \times N} = \frac{2,413606}{0,441477} = 5,47$$

Chú ý, không tính giá trị F cho ảnh hưởng mùa vụ (F_V) vì độ tự do của MS lặp lại trong vụ không đủ (df < 6).

Các giá trị F bảng tương ứng với một trong hai giá trị F tính có $df_1 = 4$ và $df_2 = 16$ là 3,01 ở mức ý nghĩa 5% và 4,77 ở mức ý nghĩa 1% (Phụ lục VI). Do đó, cả hai ảnh hưởng chính nghiệm thức và tương tác của nghiệm thức với mùa vụ có ý nghĩa cao: có sự đáp ứng về năng suất có ý nghĩa đối với nghiệm thức (áp dụng đạm) nhưng sự đáp ứng khác nhau giữa hai vụ.

Bước 6: Nếu tương tác mùa vụ x nghiệm thức (V x T) có ý nghĩa, phân chia SS tương tác thành một bộ tương phản trực giao *(orthogonal contrasts)* để có thể cung cấp thông tin về bản chất của tương tác; tại sao hiệu suất tương đối của các nghiệm thức khác nhau qua các vụ.

Trong ví dụ, nghiệm thức là các mức độ đạm, tương tác V x T tương ứng với V x N có ý nghĩa cao (Bảng 11.4). Bộ contrasts có thể giải thích bản chất của tương tác như thế là contrasts chứa các đa thức trực giao *(orthogonal polynomials)* trên đạm. Nghĩa là, $SS_{V \times N}$ nên được phân chia thành các SS của V x $N_{\text{tuyến tính}}$, V x $N_{\text{bậc hai}}$,... Vì mức độ đạm được trắc nghiệm có khoảng cách không bằng nhau, nên các hệ số đa thức trực giao nhận được cho các đa thức tuyến tính và bậc hai được chỉ dưới đây, và kết quả của sự phân chia SS được trình bày trong Bảng 11.5.

Mức độ đạm (kg/ha)	Hệ số đa thức trực giao	
	Tuyến tính	Bậc hai
0	-14	22
60	-4	-21
90	1	-21
120	6	-5
150	11	25

Bảng 11.5: Phân tích phương sai phối hợp ở Bảng 11.4 với sự phân chia SS

Nguồn biến động	Độ tự do	Tổng bình phương	Trung bình bình phương	F tính[a]
Mùa vụ (V)	1	4,495392	4,495392	[b]
Lặp lại trong vụ	4	1,261571	0,315393	
Đạm (N)	4	18,748849	4,687212	10,62**
N_L	(1)	1,435356	1,435356	$3,25^{ns}$
N_Q	(1)	17,185048	17,185048	38,93**
N_{Res}[c]	(2)	0,128445	0,064222	<1
V x N	4	9,654423	2,413606	5,47**
V x N_L	(1)	8,807778	8,087778	19,95**
V x N_Q	(1)	0,547297	0,547297	$1,24^{ns}$
V x N_{Res}	(2)	0,299348	0,149674	<1
Sai số gộp	16	7,063634	0,441477	

[a]** = có ý nghĩa ở mức 1%, [ns] = không ý nghĩa.

[b] df lặp lại trong vụ không đủ để kiểm định ý nghĩa có giá trị

[c] $N_{Res.} = N_{Residual} = N_{sai\ số}$

Phân tích chứng tỏ rằng chỉ có phần tuyến tính của hàm phản ứng thay đổi có ý nghĩa theo mùa vụ. Vì chỉ có thành phần bậc hai của SS_N có ý nghĩa, nên hàm bậc hai phù hợp với các giá trị trung bình nghiệm thức, tách biệt cho mỗi vụ, phân tích hồi qui cho kết quả:

Mùa nắng: $\hat{Y}_1 = 3,983 + 0,0523N - 0,000255N^2$ ($R^2 = 1,00$**)

Mùa mưa: $\hat{Y}_2 = 4,675 + 0,0477N - 0,000366N^2$ ($R^2 = 0,97$*)

Hai phương trình hồi qui được trình bày bằng đồ thị ở Hình 11.1. Quan sát Hình 11.1, cùng với giá trị R^2 của mỗi hồi qui đều lớn và có ý nghĩa, chứng tỏ rằng sự đáp ứng bậc hai phù hợp với số liệu một cách khá hợp lý. Mức độ gia tăng năng suất cùng với sự gia tăng mức độ đạm được áp dụng (nghĩa là, thành phần tuyến tính của hàm số) trong mùa nắng cao hơn trong mùa mưa.

Với các hàm số phản ứng của đạm được ước lượng, hai dạng mức độ đạm tối hão có thể tính là:

* Mức đạm cho năng suất tối đa:

$$N_Y = \frac{-b}{2c}$$

* Mức đạm cho lãi tối đa:

$$N_P = \frac{1}{2c}\left(\frac{P_N}{P_R} - b\right)$$

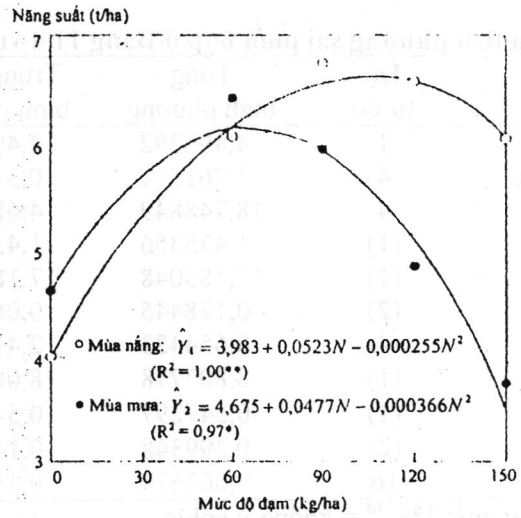
với b và c là các ước lượng của hệ số hồi qui trong phương trình $\hat{Y} = a + bN + cN^2$

P_N và P_R lần lượt là giá đạm và giá lúa.

Trong ví dụ, ước lượng các giá trị N_Y cho cả hai hồi qui là:

Mùa nắng: $N_Y = \dfrac{-0{,}0523}{(2)(-0{,}000255)} = 102{,}5$ kg N/ha

Mùa mưa: $N_Y = \dfrac{-0{,}0477}{(2)(-0{,}000366)} = 65{,}2$ kg N/ha



Hình 11.1: Ước lượng đường cong phản ứng của đạm từ số liệu ở Bảng 11.1, tách biệt cho mùa nắng và mùa mưa.

Tính giá trị N_P, chúng ta giả định tỷ lệ của giá đạm (kg/ha) trên giá lúa (t/ha) bằng 0,005. Với tỷ lệ giá này, các giá trị N_P được ước lượng là:

Mùa nắng: $N_P = \dfrac{0{,}005 - 0{,}0523}{(2)(-0{,}000255)} = 92{,}7$ kg N/ha

Mùa mưa: $N_P = \dfrac{0{,}005 - 0{,}0477}{(2)(-0{,}000366)} = 58{,}3$ kg N/ha

Kết quả thí nghiệm dường như chứng tỏ cần có sự khuyến cáo đạm khác nhau cho mùa nắng và mùa mưa.

Bước 7: Nếu kiểm định tính đồng nhất của các phương sai ở bước 4 có ý nghĩa, tiếp theo phương pháp phân chia SS tương tác V x T đã phát thảo ở bước 6, nhưng thay vì dùng MS sai số gộp như số hạng sai số để kiểm định F, SS sai số gộp đầu tiên phải được phân chia thành các thành phần tương ứng với các thành phần của $SS_{V \times T}$. Sau đó, giá trị F được tính cho mỗi thành phần của tương tác V x T bằng cách sử dụng thành phần tương ứng ở sai số gộp như là số hạng sai số của nó.

Trong ví dụ, phân tích này không cần vì kiểm định tính đồng nhất của bước 4 không ý nghĩa. Tuy nhiên, nhằm mục đích giải thích, phân tích sẽ được thực hiện bằng cách sử dụng sự phân chia $SS_{V \times N}$ giống như đã trình bày ở bước 6. Các thành phần tương ứng của SS sai số gộp được tính theo các công thức sau:

$$(\text{Lặp lại trong vụ}) \times N_L = \sum_{i=1}^{v}\left[\sum_{k=1}^{r}(SSN_L)_{ki} - (SSN_L)_i\right]$$

$$(\text{Lặp lại trong vụ}) \times N_Q = \sum_{i=1}^{v}\left[\sum_{k=1}^{r}(SSN_Q)_{ki} - (SSN_Q)_i\right]$$

với $(SSN_L)_{ki}$ là thành phần tuyến tính của SS_N được tính từ số liệu của lặp lại thứ k ở vụ thứ i. $(SSN_L)_i$ là thành phần tương ứng được tính từ giá trị tổng của tất cả lặp lại ở vụ thứ i. $(SSN_Q)_{ki}$ và $(SSN_Q)_i$ được định nghĩa tương tự nhưng cho thành phần bậc hai.

Các tổng bình phương này được tính riêng biệt cho từng vụ (Bảng 11.6).

Bảng 11.6: Tính các thành phần tuyến tính và bậc hai của SS sai số gộp trong Bảng 11.5. Số liệu của Bảng 11.1.

Số lặp lại	Tổng nghiệm thức					Tổng bình phương	
	N_0	N_1	N_2	N_3	N_4	Tuyến tính	Bậc hai
			Mùa nắng				
I	4,891	6,009	6,712	6,458	5,683	0,646228	1,227907
II	2,577	6,625	6,693	6,675	6,868	9,636871	3,555132
III	4,541	5,672	6,799	6,636	5,692	1,425382	1,386474
Tổng	12,009	18,306	20,204	19,769	18,243	8,677167	5,799363
			Mùa mưa				
I	4,999	6,351	6,071	4,818	3,436	1,382265	3,935604
II	3,503	6,316	5,969	4,024	4,047	0,000284	4,964835
III	5,356	6,582	5,893	5,813	3,740	1,017294	3,151471
Tổng	13,858	19,249	17,933	14,655	11,223	1,565967	11,932982

Ví dụ, thành phần tuyến tính của SS_N cho lặp lại I của mùa nắng được tính như sau:

$$(SSN_L)_{11} = \frac{[-14(4,891) - 4(6,009) + (6,712) + 6(6,458) + 11(5,683)]^2}{(-14)^2 + (-4)^2 + (1)^2 + (6)^2 + (11)^2}$$

$$= \frac{(15,463)^2}{370} = 0,646228$$

Sau cùng, các thành phần SS sai số gộp được tính như sau:

$$(\text{Lặp lại trong vụ}) \times N_L = [(0,646228 + 9,636871 + 1,425382) - 8,677167] + [(1,382265 + 0,000284 + 1,017294) - 1,565967]$$
$$= 3,031314 + 0,833876 = 3,865190$$

$$(\text{Lặp lại trong vụ}) \times N_Q = [(1,227907 + 3,555132 + 1,386474) - 5,799363] + [(3,935604 + 4,946835 + 3,151471) - 11,932982]$$
$$= 0,370150 + 0,100928 = 0,471078$$

$$(\text{Lặp lại trong vụ}) \times N_{Res.} = \text{SS sai số gộp} - (\text{Lặp lại trong vụ}) \times SSN_L - (\text{Lặp lại trong vụ}) \times SSN_Q$$
$$= 7,063634 - 3,865190 - 0,471078 = 2,727366$$

Các kết quả này được tóm tắt như sau:

Nguồn biến động	Độ tự do	Tổng bình phương	Trung bình bình phương
Sai số gộp	(16)	(7,063634)	(0,441477)
(Lặp lại trong vụ) x N_L	4	3,865190	0,966298
(Lặp lại trong vụ) x N_Q	4	0,471078	0,117770
(Lặp lại trong vụ) x $N_{Res.}$	8	2,727366	0,340921

Ba thành phần phương sai gộp sau đó được sử dụng để kiểm định ý nghĩa của các thành phần tương ứng của tương tác V x N trong Bảng 11.5 như sau:

$$F(VxN_L) = \frac{MS\ V \times N_L}{(\text{Lặp lại trong vụ}) \times MSN_L} = \frac{8,807778}{0,966298} = 9,11$$

$$F(V \times N_Q) = \frac{MS\ V \times N_Q}{(\text{Lặp lại trong vụ}) \times MSN_Q} = \frac{0,547297}{0,117770} = 4,65^a$$

$$F(VxN_{Res.}) = \frac{MS\ V \times N_{Res.}}{(\text{Lặp lại trong vụ}) \times MSN_{Res.}} = \frac{0,149674}{0,340921} < 1^a$$

[a]Mặc dù độ tự do của MS (Lặp lại trong vụ) x N_L và MS (Lặp lại trong vụ) x N_Q không đủ để kiểm định ý nghĩa có giá trị, nhưng với ý định giải thích, có thể lờ đi sự thiếu hụt như thế.

2. Phân tích qua nhiều năm *(Analysis over years)*

Trái với mùa trồng trong năm, sự biến thiên của môi trường qua các năm thường không thể đoán trước được; chẳng hạn, không lý do gì để cho rằng năm có số lẻ khô hơn năm có số chẵn hoặc cũng không thể đoán trước một cách chính xác năm nào hạn hán hay lũ lụt hoặc mức độ thiệt hại do dịch sâu bệnh,... Vì vắng mặt mô hình có thể dự đoán, nên nhìn chung các năm được xem như *biến ngẫu nhiên*: một kỹ thuật tốt phải tỏ ra tốt hơn chắc chắn qua nhiều năm cho ít nhất một mùa trồng nào đó. Vì vậy, mục đích đầu tiên của phân tích phối hợp của số liệu qua nhiều năm là để nhận ra các kỹ thuật mà tác động trung bình của chúng qua nhiều năm luôn cao và ổn định. Do đó, tương tác giữa nghiệm thức và năm không có ý nghĩa nông học và ít quan trọng hơn tương tác giữa nghiệm thức và mùa vụ.

Phương pháp phối hợp số liệu qua nhiều năm được giải thích với thí nghiệm gồm bảy giống lúa được trắc nghiệm qua hai năm, bằng cách sử dụng bố trí RCBD với ba lần lặp lại. Năng suất hạt được ghi nhận trong Bảng 11.7, phân tích phương sai của từng thí nghiệm trong Bảng 11.8 và các trung bình giống cho từng năm trong Bảng 11.9. Lưu ý, để phân tích phối hợp các số liệu qua nhiều năm cần thỏa điều kiện các thí nghiệm chỉ thực hiện cùng vụ trồng (hoặc vụ nắng hoặc vụ mưa) để tác động trung bình qua các năm cao và ổn định.

Các bước phân tích phối hợp được thực hiện như sau:

Bước 1: Phát thảo phân tích phương sai phối hợp qua các năm, dựa trên bố trí RCBD được trình bày trong Bảng 11.10. Chú ý sự khác nhau đầu tiên giữa phân tích phương sai phối hợp qua các năm và qua các mùa vụ là ở kiểm định F để kiểm định ý nghĩa của ảnh hưởng nghiệm thức. Khi số liệu được phối hợp qua nhiều vụ (nghĩa là biến cố định) số hạng sai số là MS sai số gộp; còn khi phối hợp qua nhiều năm (nghĩa là biến ngẫu nhiên) số hạng sai số là MS tương tác năm x nghiệm thức.

Bảng 11.7: Năng suất hạt (t/ha) của bảy giống lúa được trắc nghiệm, dùng bố trí RCBD với ba lần lặp lại, trong cùng mùa vụ qua hai năm. (Trích từ Bảng 8.7 của Gomez & Gomez. 19834. Trang 328)

Số giống	Năm 1				Năm 2			
	Llại I	Llại II	Llại III	Tổng	Llại I	Llại II	Llại III	Tổng
1	3,036	4,177	3,884	11,097	1,981	3,198	3,726	8,905
2	1,369	1,554	1,899	4,822	3,751	2,391	3,714	9,856
3	5,311	5,091	4,839	15,241	3,868	3,134	3,487	10,489
4	2,559	3,980	3,853	10,392	2,729	2,786	2,598	8,113
5	1,291	1,705	2,130	5,126	3,222	3,554	2,452	9,228
6	3,452	3,548	4,640	11,640	4,250	4,134	3,339	11,723
7	1,812	2,914	0,958	5,684	3,336	4,073	2,885	10,294
Tổng				64,002				68,608

Bảng 11.8: Phân tích phương sai qua từng năm của thí nghiệm giống lúa (RCBD) với bảy giống và ba lần lặp lại. Tính từ số liệu của Bảng 11.7.

Nguồn biến động	Độ tự do	Tổng bình Phương	Trung bình bình phương	F (tính)	F bảng 5%	F bảng 1%
		Năm 1				
Lặp lại	2	1,38549	0,69274			
Giống	6	31,85691	5,30948	16,11**	3,00	4,82
Sai số	12	3,95588	0,32966			
		Năm 2				
Lặp lại	2	0,09698	0,04849			
Giống	6	2,79800	0,46633	$1,16^{ns}$	3,00	4,82
Sai số	12	4,84346	0,40362			

* = Khác biệt ở mức ý nghĩa 5%

** = Khác biệt ở mức ý nghĩa 1%

Bảng 11.9: Năng suất trung bình của bảy giống lúa được trắc nghiệm qua hai năm liên tục. Tính từ số liệu của Bảng 11.7.

Số giống	Năng suất trung bình (t/ha)			Sai biệt
	Năm 1	Năm 2	Trung bình	
1	3,699	2,968	3,334	- 0,731
2	1,607	3,285	2,446	1,678
3	5,080	3,496	4,288	- 1,584
4	3,464	2,704	3,084	- 0,760
5	1,709	3,076	2,392	1,367
6	3,880	3,908	3,894	0,028
7	1,895	3,431	2,663	1,536
Tổng	3,048	3,267	3,158	0,219

Bước 2: Tính các tổng bình phương (SS) và trung bình bình phương (MS) tương tự như phương pháp phân tích qua nhiều vụ, bằng cách thay thế mùa vụ bằng năm. Trong ví dụ, các SS và MS được tính như sau:

SS lặp lại trong năm = 1,38549 + 0,09698 = 1,48247

SS sai số gộp = 3,95588 + 4,84346 = 8,79934

$$C.F. = \frac{\left(\sum_{i=1}^{y} G_i\right)^2}{yrv} = \frac{(64{,}002 + 68{,}608)^2}{42} = 418{,}70029$$

$$SS_Y = \frac{(64{,}002)^2 + (68{,}608)^2}{(3).(7)} - 418{,}70029 = 0{,}50512$$

$$SS_V = \frac{(10{,}002 + 14{,}678)^2 + \cdots + (15{,}978)^2}{(3)(2)} - 418{,}70029 = 19{,}15891$$

$$SS_{YxV} = \frac{(11{,}097)^2 + (4{,}822)^2 + \cdots + (10{,}294)^2}{3} - 418{,}70029 - 0{,}50512 - 19{,}15891$$

$$= 15{,}49600$$

Bảng 11.10: Phát thảo phân tích phương sai phối hợp qua nhiều năm[a], dựa trên bố trí RCBD với t nghiệm thức và r lần lặp lại.

Nguồn biến động	Độ tự do*	Trung bình bình phương	F tính
Năm	y – 1	MS_Y	$\frac{MS_Y}{MS_R}$
Lặp lại trong năm	y(r – 1)	MS_R	
Nghiệm thức (V)	v -1	MS_V	$\frac{MS_V}{MS_{YxV}}$
Y x V	(y – 1)(v – 1)	MS_{YxV}	$\frac{MS_{YxV}}{MS_E}$
Sai số gộp	y(r -1)(v -1)	MS_E	
Tổng cộng	yrv - 1		

[a]Năm được khảo sát như biến ngẫu nhiên.
* y = số năm, v = số giống và r = số lần lặp lại

Các giá trị MS được tính theo cách thông thường. Kết quả được trình bày trong Bảng 11.11.

Bảng 11.11: Phân tích phương sai phối hợp của các thí nghiệm RCBD qua hai năm. Tính từ số liệu ở Bảng 11.7 và 11.8.

Nguồn biến động	Độ tự do	Tổng bình phương	Trung bình bình phương	F (tính)	F bảng 5%	F bảng 1%
Năm (Y)	1	0,50512	0,50512	[b]		
Lặp lại trong năm	4	1,48247	0,37062			
Giống (V)	6	19,15891	3,19315	1,24[ns]	4,28	8,47
Y x V	6	15,49600	2,58267	7,04**	2,51	3,67
Sai số gộp	24	8,79934	0,36664			
Tổng cộng	41	45,44184				

** = Ý nghĩa ở mức 1%, [ns] = Không ý nghĩa
[b]df của lặp lại trong năm không đủ để kiểm định ý nghĩa có giá trị.

Bước 3: Kiểm định tính đồng nhất của MS sai số từ phân tích phương sai của từng thí nghiệm bằng cách dùng kiểm định F hoặc kiểm định chi bình phương.

Trong ví dụ, áp dụng kiểm định F:

$$F = \frac{\text{MSsai số lớn}}{\text{MSsai số nhỏ}} = \frac{0,40362}{0,32966} = 1,22$$

Vì giá trị F tính nhỏ hơn F bảng là 2,69 với df_1 (tử số) = df_2 (mẫu số) = 12 ở mức ý nghĩa 5%. Do đó, không bác bỏ giả thuyết về tính đồng nhất của các phương sai sai số.

Bước 4: Nếu các phương sai sai số không đồng nhất ở bước 3, chuyển đến bước 5. Nếu chúng đồng nhất, tính các giá trị F để kiểm định ý nghĩa của các ảnh hưởng khác như chỉ ở Bảng 11.10.

Trong ví dụ, vì kiểm định F về tính đồng nhất của các phương sai sai số ở bước 4 cho kết quả không ý nghĩa, nên tính các giá trị F theo các công thức ở Bảng 11.10. Do độ tự do của SS lặp lại trong năm chỉ bằng 4, nên không tính giá trị F để kiểm định ảnh hưởng chính của năm. Các giá trị F khác được tính như sau:

$$F_V = \frac{MS_V}{MS_{YxV}} = \frac{3,19315}{2,58267} = 1,24$$

$$F_{YxV} = \frac{MS_{YxV}}{\text{MS sai số gộp}} = \frac{2,58267}{0,36664} = 7,04$$

So sánh các giá trị F này với các giá trị F bảng tương ứng, theo cách thông thường, cho thấy F_V không ý nghĩa, và F_{YxV} có ý nghĩa ở mức 1%; do đó, ảnh hưởng tương tác giữa giống và năm có ý nghĩa cao, nhưng ảnh hưởng trung bình của giống không ý nghĩa (Bảng 11.11).

Bước 5: Nếu kiểm định tính đồng nhất của các phương sai ở bước 4 có ý nghĩa, chia SS sai số gộp dựa vào bộ contrasts mong muốn trên $SS_{Y\,x\,V}$.

Trong ví dụ, vì các phương sai sai số đồng nhất, nên không cần thiết chia SS sai số gộp. Nghĩa là, ảnh hưởng Y x V, cũng như bất cứ thành phần nào của nó, có thể được kiểm định bằng cách sử dụng MS sai số gộp như số hạng sai số.

Bước 6: Nếu tương tác nghiệm thức x năm có ý nghĩa, kiểm tra độ lớn của nó có liên quan đến ảnh hưởng trung bình của các nghiệm thức. Nếu ảnh hưởng tương tác nhỏ có liên quan đến ảnh hưởng trung bình, thì việc xếp hạng các nghiệm thức qua các năm mong đợi sẽ ổn định (nghĩa là, nghiệm thức A cho năng suất cao hơn nghiệm thức B ở tất cả các năm mặc dù trong một vài năm sự khác biệt hơi cao hơn ở những năm khác) và tương tác có thể được lờ đi.

Mặt khác, nếu ảnh hưởng tương tác tương đối lớn và sự xếp hạng các nghiệm thức thay đổi qua các năm (nghĩa là, lúc nghiệm thức A cho năng suất tốt hơn nghiệm thức B trong một vài năm, và không tốt hơn hoặc thậm chí xấu hơn ở các năm khác).

Trong ví dụ, tương tác quá lớn đến nỗi ảnh hưởng trung bình của giống không có ý nghĩa. Qua sự kiểm tra sự khác biệt trung bình giữa các năm đối với một trong bảy giống (Bảng 11.9, cột cuối) có thể nhận ra ba nhóm giống. Nhóm đầu tiên gồm các giống 2, 5 và 7 có năng suất ở năm 2 cao hơn năm 1; nhóm thứ hai bao gồm các giống 1, 3 và 4 cho năng ở năm 1 cao hơn năm 2; và nhóm thứ ba gồm có giống 6 cho năng suất như nhau ở cả hai năm.

Mặc dù rõ ràng trong thí nghiệm này không thể nhận ra giống tốt một cách chắc chắn, nhưng cần lưu ý cho các thí nghiệm sau này có lẽ nên dành cho giống 6, là giống luôn luôn biểu hiện cao. Đối với giống 3, mặc dù cho năng suất trung bình cao nhất (4,288 t/ha) qua các năm, nhưng chỉ cho năng suất cực cao ở năm đầu tiên (5,08 t/ha).

3. Phân tích thí nghiệm qua nhiều địa điểm *(Analysis over sites)*

Phương pháp để phối hợp số liệu từ thí nghiệm tính thích nghi của giống ở m địa điểm được trình bày dưới đây cho thí nghiệm bao gồm bảy giống lúa được trắc nghiệm theo kiểu bố trí RCBD với ba lần lặp lại ở chín địa điểm. Kết quả phân tích phương sai ở từng địa điểm, dựa trên số liệu năng suất hạt được ghi nhận ở Bảng 11.12. Năng suất trung bình của giống ở mỗi địa điểm được trình bày trong Bảng 11.13. Các bước thực hiện như sau:

Bước 1: Áp dụng kiểm định chi bình phương cho tính đồng nhất của phương sai đối với m phương sai sai số của phân tích phương sai ở từng địa điểm. Nếu kiểm định có ý nghĩa, tất cả địa điểm có hệ số biến động cực lớn (nghĩa là, cv > 20%) có thể được loại ra khỏi phân tích phối hợp.

Trong ví dụ, kiểm định χ^2 được áp dụng đối với chín trung bình bình phương sai số ở Bảng 11.12 cho kết quả giá trị χ^2 bằng 59,65, có ý nghĩa ở mức 1%. Tuy nhiên, vì không có địa điểm nào có giá trị cv lớn hơn 20%, nên tất cả chín địa điểm đều được đưa vào phân tích phối hợp.

Bước 2: Phát thảo phân tích phối hợp trên m địa điểm và tính SS và MS khác nhau, theo phương pháp đã được phát thảo cho số liệu phối hợp qua các năm, với năm được thay thế bằng địa điểm (L). Trong ví dụ này, SS được tính dựa trên các giá trị trung bình ở Bảng 11.13 và tổng bình phương ở Bảng 11.12 như sau:

SS Lặp lại trong địa điểm = 0,01581 + 1,36015 + . . . + 0,08317 = 7,03960

SS sai số gộp = 3,25070 + 4,25357 + . . . + 0,52724 = 44,81283

$C.F. = (3).(7).(9).(4,688)^2 = 4153,71802$

$SS_L = (3)(7).[(4,795)^2 + \cdots + (4,729)^2] - 4153,71802 = 647,87421$

$SS_V = (3).(9).[(5,000)^2 + \cdots + (4,708)^2] - 4153,71802 = 8,85162$

$SS_{VxL} = (3).[(4,835)^2 + \cdots + (4,224)^2] - 4153,71802 - 647,87421 - 8,85162 = 45,63398$

Các giá trị MS được tính theo cách thông thường. Kết quả được trình bày trong Bảng 11.14.

Bước 3: Nhận dạng khoảng thích nghi của các nghiệm thức qua việc phân chia thích hợp tổng bình phương tương tác nghiệm thức x địa điểm. Điều này có thể được thực hiện bởi một hoặc phối hợp các phương pháp sau đây:

* Phương pháp địa điểm đồng nhất: Sắp xếp các địa điểm thành các nhóm đồng nhất để tương tác giữa nghiệm thức và địa điểm trong mỗi nhóm không ý nghĩa. Sau đó có thể nhận dạng một bộ giống riêng biệt như các giống thích nghi để khuyến cáo mỗi bộ cho các địa điểm đồng nhất.

* Phương pháp nghiệm thức đồng nhất: Nhận dạng một hoặc nhiều nghiệm thức mà sự đóng góp của chúng vào tương tác giữa nghiệm thức và địa điểm cao. Các nghiệm thức có năng suất dao động lớn qua các địa điểm là những nghiệm thức đóng góp chủ yếu vào tương tác nghiệm thức x địa điểm và được dự đoán có khoảng thích nghi hẹp.

Để quyết định áp dụng phương pháp nào, cần khảo sát số liệu bằng cách (1) vẽ đồ thị các giá trị trung bình của mỗi giống ở trục Y đối với ký hiệu địa điểm (nghĩa là, L_1, L_2, L_3,...) ở trục X, (2) vẽ đồ thị các giá trị trung bình của mỗi địa điểm ở trục Y đối với ký hiệu giống ở trục X, hoặc (3) vẽ đồ thị cả hai. Nếu đồ thị cách (1) phát hiện một hoặc nhiều giống khác nhau rõ rệt, đề nghị áp dụng phương pháp nghiệm thức đồng nhất. Tuy nhiên, nếu quan sát thấy dạng đồ thị (2) chứng tỏ sự gom nhóm các địa điểm rõ ràng đối với phản ứng giống nhau của giống, đề nghị áp dụng phương pháp địa điểm đồng nhất.

Bảng 11.12: Phân tích phương sai của từng thí nghiệm (RCBD). Một trong chín địa điểm (L1 đến L9). Số liệu của thí nghiệm về giống với bảy giống lúa.

N.B.Đ.	ĐTD	L_1	L_2	L_3	L_4	L_5	L_6	L_7	L_8	L_9
Lặp lại	2	0,01581	1,36015	0,26245	0,13993	0,03304	0,09698	4,33975	0,70832	0,08317
Giống	6	6,75774	5,91469^{ns}	7,44175^{ns}	8,60013**	8,34556**	2,89800^{ns}	7,76741^{ns}	3,07978^{ns}	3,49240**
Sai số	12	3,25070	4,25357	20,58217	1,33921	1,51767	4,84346	5,33916	4,15965	0,52724
cv (%)		10,9	16,2	17,0	6,5	14,7	19,4	9,0	14,5	4,4

** = Kiểm định F có ý nghĩa ở mức 5%

* = Kiểm định F có ý nghĩa ở mức 1%

^{ns} = Kiểm định F không ý nghĩa

Bảng 11.13: Năng suất trung bình (t/ha) của bảy giống lúa (V_1 đến V_7) được trắc nghiệm ở chín địa điểm (L_1 đến L_9)[a].

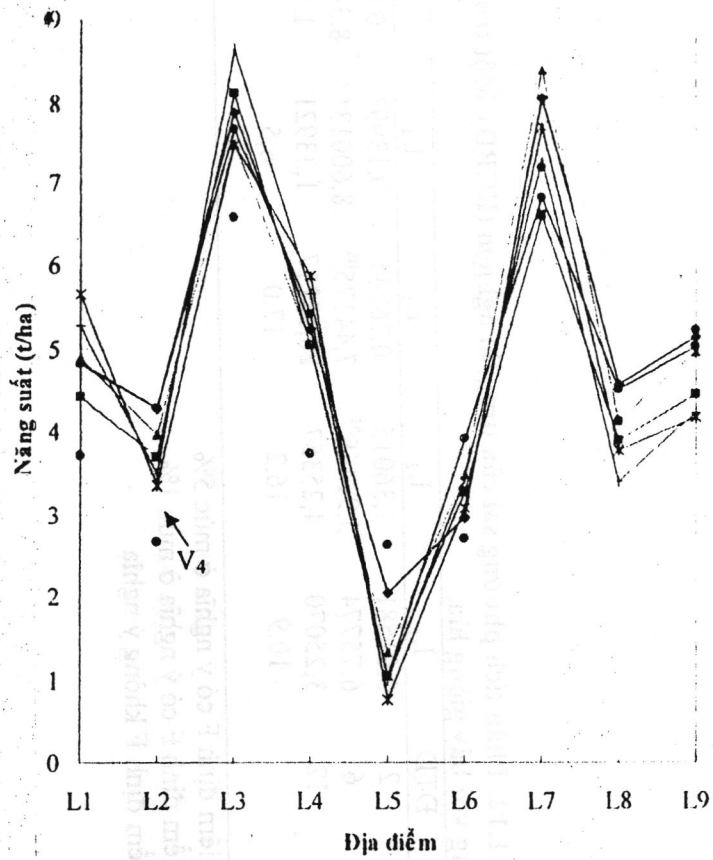
Giống	Địa điểm									Trung bình
	L_1	L_2	L_3	L_4	L_5	L_6	L_7	L_8	L_9	
V_1	4,835	4,288	7,882	5,219	2,052	2,968	8,066	4,568	5,122	5,000
V_2	4,412	3,694	8,110	5,036	1,042	3,285	6,584	3,889	4,445	4,500
V_3	4,888	3,963	7,495	5,054	1,342	3,496	8,387	4,136	4,945	4,856
V_4	3,717	2,675	6,568	3,725	2,643	2,704	7,182	4,136	5,200	4,283
V_5	5,635	3,351	7,475	5,858	0,772	3,076	7,679	3,765	4,159	4,641
V_6	4,808	4,288	7,672	5,403	1,042	3,908	6,796	4,506	5,012	4,826
V_7	5,271	3,495	8,652	5,645	1,000	3,431	7,261	3,395	4,224	4,708
T.B.	4,795	3,679	7,693	5,134	1,413	3,267	7,422	4,056	4,730	4,688

[a]Phân tích phương sai cho từng địa điểm ở Bảng 11.12

Trong ví dụ, đồ thị (1) ở Hình 11.2 cho thấy V_4 có khuynh hướng khác rõ rệt với những giống còn lại. Do đó, đề nghị sử dụng phương pháp nghiệm thức đồng nhất và nên phân chia SS_V và SS_{VxL} thành các thành phần sau:

E_1 : V_4 với các giống khác

E_2 : giữa các giống khác



Hình 11.2: Xác định tương tác giống x địa điểm bằng cách vẽ đồ thị năng suất trung bình mỗi giống theo địa điểm của số liệu ở Bảng 11.11 (Trích dẫn từ Hình 8.2 của Gomez and Gomez. 1984. Trang 337).

Bảng 11.14: Phân tích phương sai phối hợp sơ bộ qua các địa điểm được tính từ số liệu ở Bảng 11.12 và 11.13.

Nguồn biến động	Độ tự do	Tổng bình phương	Trung bình bình phương
Địa điểm (L)	8	647,87421	80,98428
Lặp lại trong địa điểm	18	7,03960	0,39109
Giống (V)	6	8,85162	1,47527
V x L	48	45,63398	0,95071
Sai số gộp	108	44,81283	0,41493
Tổng cộng	188	754,21224	

Kết quả của sự phân chia này được trình bày trong Bảng 11.15

Bước 4: Tính F để kiểm định ý nghĩa của SS_V và SS_{VxL} và các thành phần của chúng tùy thuộc vào mức ý nghĩa của kiểm định tính đồng nhất của phương sai sai số qua các địa điểm được tính ở bước 1.

Trong ví dụ, kiểm định tính đồng nhất của các phương sai sai số ở bước 1 có ý nghĩa. Vì thế, thực hiện việc phân chia tổng bình phương sai số gộp tương ứng với bộ contrast mong muốn trên các giống nhận được từ bước 3. Kết quả được trình bày ở Bảng 11.15.

Bảng 11.15: Phân chia tổng bình phương trong phân tích phương sai phối hợp sơ bộ ở Bảng 11.14, để đối phó với tính không đồng nhất của các phương sai sai số qua các địa điểm.

Nguồn biến động	Độ tự do	Tổng bình phương	Trung bình bình phương	F (tính)	F (bảng)	
Địa điểm (L)	8	647,87421	80,98428	207,01**	2,51	3,71
Lặp lại trong địa điểm	18	7,03960	0,39109			
Giống (V)	6	8,85162	1,47527	3,56**	2,18	2,56
E1: V4 với các giống khác	(1)	5,15950	5,15950	12,43**	3,93	6,88
E2: giữa các giống khác	(5)	3,69212	0,73842	$1,78^{ns}$	2,30	3,19
V x L	48	45,63398	0,95071	2,29**	1,48	1,76
E1 x L	(8)	21,22237	2,65280	6,39**	2,02	2,68
E2 x L	(40)	24,41161	0,61029	$1,47^{ns}$	1,50	1,78
Sai số gộp	108	44,81283	0,41493			
(lặp lại trong địa điểm) x E1	(18)	5,05481	0,28082			
(lặp lại trong địa điểm) x E2	(90)	39,75802	0,44176			

** = có ý nghĩa ở mức 1%, ^{ns} = không ý nghĩa

Kiểm định F đối với tính đồng nhất của hai thành phần SS sai số gộp cho giá trị F tính là 1,57; không có ý nghĩa. Do đó, MS sai số gộp có thể được dùng làm mẫu số của các giá trị F để kiểm định ý nghĩa ảnh hưởng của giống (V), hoặc ảnh hưởng tương tác giống x địa điểm (V x L), hoặc các thành phần của chúng. Ví dụ,

$$F_V = \frac{MS_V}{MSsai\ số\ gộp} = \frac{1,47527}{0,441403} = 3,36$$

có ý nghĩa ở mức 1%. Kết quả của tất cả các kiểm định được trình bày trong Bảng 11.15 cho thấy, ngoại trừ giống V4, các giống còn lại khác biệt không ý nghĩa và kết quả này giống nhau ở tất cả các địa điểm. Điều này muốn nói rằng ngoài giống V_4, tất cả các giống khác đều thích nghi như nhau ở chín địa điểm. Tuy nhiên, năng suất của giống V_4 (Bảng 11.13) dao động đáng kể qua các địa điểm chứng tỏ giống này thích

nghi hẹp. Hơn nữa, vì năng suất trung bình của giống này cũng thấp hơn các giống khác có ý nghĩa, nên tiềm năng của nó để khuyến cáo, ngay cả ở một vài địa điểm cho năng suất tương đối cao, cũng còn nghi ngờ.

II. BIẾN ĐỔI SỐ LIỆU *(Data transformation)*

Cách sửa chữa tạm thời khi các tiền đề cơ bản của phân tích phương sai không thoả là thực hiện việc biến đổi số liệu. Nó tương đương với sự thay đổi thang số đo của một hiện tượng, với mục đích cải thiện các đặc tính của số liệu mà có thể làm vô hiệu ít nhiều trong phân tích phương sai.

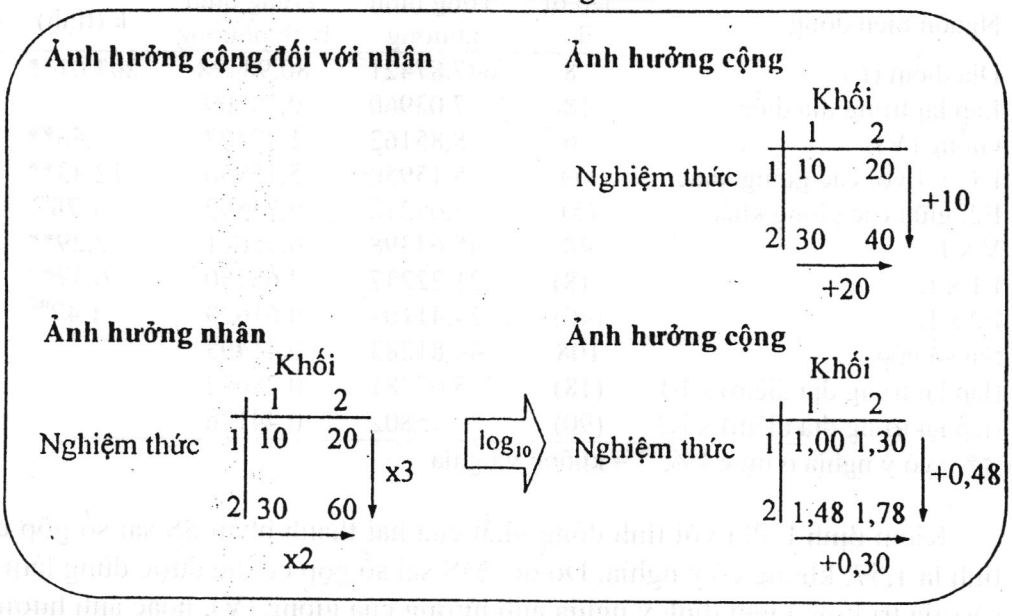
Các ảnh hưởng không cộng tính, sai số thí nghiệm không phân bố chuẩn chuẩn và phương sai sai số không đồng nhất là ba điều kiện để áp dụng việc biến đổi số liệu. Trái lại, không biến đổi số liệu không thể đền bù cho các sai số không độc lập.

Với phương pháp biến đổi số liệu, các số liệu gốc sẽ được biến đổi thành bộ số liệu mới với hy vọng thỏa được điều kiện về tính đồng nhất của phương sai. Do sự biến đổi được áp dụng cho tất cả các số liệu quan sát, nên các so sánh giữa các nghiệm thức không bị thay đổi và vẫn giữ giá trị của chúng. Có ba cách biến đổi số liệu.

1. Đổi sang logarith thập phân

Biến đổi logarith cho phép cải thiện tính đồng nhất của phương sai và ảnh hưởng cộng tính. Nó được áp dụng khi các nghiệm thức (nguồn biến động) có ảnh hưởng nhân thay vì cộng (nhân được thay thế bằng cộng: A x B được thay bằng log A + log B).

Biến đổi logarith



Biến đổi logarith thích hợp nhất đối với các số liệu có độ lệch chuẩn tỷ lệ với trung bình. Thường các điều kiện này gặp ở số liệu là những con số nguyên và giá trị của chúng có khoảng biến động khá lớn. Chẳng hạn, số liệu về số côn trùng trên một lô hoặc số ổ trứng trên một cây (hoặc trên một đơn vị diện tích)

Để đổi số liệu sang logarith, chúng ta lấy log của tất cả số liệu quan sát. Nếu bộ số liệu có các giá trị nhỏ (< 10), nên dùng log(X + 1) thay vì log X (X là số liệu gốc) vì không có logarith của 0 và logarith của tỷ lệ giữa 0 và 1 có giá trị âm.

Ví dụ: Số liệu về số ấu trùng sống sót trên cây lúa được xử lý với các tỷ lệ thuốc trừ sâu khác nhau. Thí nghiệm được bố trí theo thể thức khối hoàn toàn ngẫu nhiên (RCBD), bốn lần lặp lại (Bảng 11.16).

Do một vài giá trị ở Bảng 11.16 nhỏ hơn 10, nên áp dụng log(X + 1) thay vì log X. Tính log(X + 1) cho tất cả số liệu, với X là số liệu gốc ở Bảng 11.16. Số liệu biến đổi được trình bày trong Bảng 11.17.

Phân tích phương sai theo cách thông thường dựa trên số liệu biến đổi ở Bảng 11.17. Kết quả được ghi nhận ở Bảng 11.18.

Kiểm định đối với các so sánh cặp (LSD, Duncan,...) phải được thực hiện trên các trung bình biến đổi. Tuy nhiên, trình bày các trung bình nghiệm thức sau cùng dựa trên số liệu gốc.

Ví dụ: Dùng kiểm định Duncan để so sánh tất cả các cặp trung bình dựa trên trung bình biến đổi. Kết quả được trình bày ở Bảng 11.19.

Sau đó tính trung bình nghiệm thức dựa trên số liệu gốc ở Bảng 11.16 hoặc chuyển trung bình biến đổi ở Bảng 11.17 sang tỷ lệ gốc (nghĩa là, tính antilog của mỗi trung bình biến đổi và trừ đi 1). Theo lý thuyết, phương pháp chuyển các trung bình biến đổi sang antilog thích hợp hơn, nhưng trong thực hành, để đơn giản, người ta thường tính trung bình nghiệm thức từ số liệu gốc.

Mặc dầu có thể tính trung bình nghiệm thức từ số liệu gốc, nhưng phải bảo đảm số thứ hạng của các trung bình gốc và trung bình biến đổi không quá khác nhau làm ảnh hưởng đến thứ tự so sánh trung bình. Một vài trường hợp hiếm, điều kiện này không thỏa như giải thích ở Bảng 11.20, trong đó trung bình gốc của nghiệm thức thứ 5 cao hơn nghiệm thức 4 và 6, nhưng thấp hơn trung bình biến đổi. Trong trường hợp như thế, phương pháp chuyển trung bình biến đổi sang antilog sẽ thích hợp hơn.

Trong ví dụ, phương pháp đầu tiên tính trung bình nghiệm thức từ số liệu gốc (Bảng 11.16) và kết quả được trình bày sau cùng ở Bảng 11.19. Kết quả Duncan được áp dụng trước tiên với trung bình biến đổi, sau đó chuyển trực tiếp sang trung bình gốc (xem cột cuối của Bảng 11.19).

Bảng 11.16: Số ấu trùng còn sống được tìm thấy ở các nghiệm thức xử lý thuốc trừ sâu khác nhau.
(Trích từ Bảng 7.14 của Gomez & Gomez. 1984. Trang 300)

STT	Nghiệm thức	Số ấu trùng				Tổng N.T.	T. bình N.T.
		Llại I	Llại II	Llại III	Llại IV		
1	Diazinon (4)	9	12	0	1	22	5,50
2	Diazinon (3)	4	8	5	1	18	4,50
3	Diazinon (2)	6	15	6	2	29	7,25
4	Diazinon (1)	9	6	4	5	24	6,00
5	Diazinon (2) + MLVC (2)	27	17	10	10	64	16,00
6	Diazinon (2) + MLVC + SLVC (2)	35	28	2	15	80	20,00
7	Diazinon (1) ở giai đoạn nhiễm 15%	1	0	0	0	1	0,25
8	Diazinon (1) ở giai đoạn nhiễm 20%	10	0	2	1	13	3,25
9	Đối chứng	4	10	15	5	34	8,50
T.cộng		105	96	44	40	285	

* Số trong ngoặc chỉ số lần áp dụng hóa chất

Bảng 11.17: Đổi số liệu ở Bảng 10.16 sang logarith, log(X + 1)

Số nghiệm thức	Lặp lại I	II	III	IV	Tổng (T)	Trung. bình	Antilog của X
1	1,0000	1,1139	0,0000	0,3010	2,4149	0,6037	4,02
2	0,6990	0,9542	0,7782	0,3010	2,7324	0,6831	4,82
3	0,8451	1,2041	0,8451	0,4771	3,3714	0,8428	6,96
4	1,0000	0,8451	0,6990	0,7782	3,3223	0,8306	6,77
5	1,4472	1,2553	1,0414	1,0414	4,7853	1,1963	15,71
6	1,5563	1,4624	1,4771	1,2041	4,6999	1,1750	14,96
7	0,3010	0,0000	0,0000	0,0000	0,3010	0,0752	1,19
8	1,0414	0,0000	0,4771	0,3010	1,8195	0,4549	2,85
9	0,6990	1,0414	1,2041	0,7782	3,7227	0,9307	8,53
Tổng	8,5890	7,8764	5,5220	5,1820	27,1694		

Bảng 11.18: Phân tích phương sai của số liệu ở Bảng 10.17

Nguồn biến động	Độ tự do	Tổng bình phương	Trung bình bình phương	F (tính)
Lặp lại	3	0,95666	0,31889	
Nghiệm thức	8	3,98235	0,49779	5,70**
Sai số	24	2,09615	0,08734	
Tổng cộng	35	7,03516		

** = Khác biệt ở mức ý nghĩa 1%

Bảng 11.19: Áp dụng kiểm định Duncan để so sánh các trung bình nghiệm thức ở Bảng 11.16 qua so sánh DUNCAN của các trung bình đã biến đổi ở Bảng 11.17.

Số nghiệm thức	Bước sơ bộ: Trung bình đã đổi	Duncan	Trình bày sau cùng: Trung bình gốc	Duncan[a]
1	0,6037	b	5,50	b
2	0,6831	b	4,50	b
3	0,8428	bc	7,25	bc
4	0,8306	bc	6,00	bc
5	1,1963	c	16,00	c
6	1,1750	c	20,00	c
7	0,0752	a	0,25	a
8	0,4549	ab	3,25	ab
9	0,9307	bc	8,50	bc

[a]Các trung bình có cùng chữ theo sau thì không khác biệt ý nghĩa ở mức 5%.

2. Đổi sang căn bậc hai *(Square-Root Transformation)*

- Áp dụng khi số liệu có dạng phân bố Poisson. Luật Poisson mô tả phân bố của các biến cố hiếm xuất hiện thình lình trong không gian hoặc thời gian. Do đó, nó thích hợp đối với các số liệu đếm (số nguyên nhỏ) và không phải là số liệu đo lường. Chẳng hạn, số liệu đạt được trong việc đếm các biến cố hiếm như số cây bị nhiễm bệnh trên lô, số côn trùng bắt được trong một bẫy, số cỏ dại trên lô, số vi khuẩn trong đĩa petri,...

- Khi số liệu phân bố theo luật Poisson, lúc đó chứng tỏ các kết quả đếm được lặp lại trên các đơn vị thí nghiệm có phân bố quanh trung bình với phương sai bằng chính số trung bình (độ lệch chuẩn bằng căn bậc hai của trung bình). Trong trường hợp này, phương sai có khuynh hướng tỷ lệ với trung bình. Để thực hiện phân tích phương sai

cho các số liệu như thế phải biến đổi sang căn bậc hai để trả lại phương sai độc lập với trung bình.

Bảng 11.20: Giải thích sự khác nhau về thứ tự xếp hạng của trung bình nghiệm thức dựa trên tỷ lệ liệu gốc và tỷ lệ đã biến đổi[a].

Số nghiệm thức	Số rầy xanh/lồng		Đổi sang logarith		Antilog của trung bình đã đổi
	Trung bình	Hạng	Trung bình	Hạng	
1	73,0	4	1,75872	4	57,4
2	22,8	10	1,34197	7	22,0
3	112,3	1	2,01669	1	103,9
4	25,5	9	1,26622	9	18,5
5	36,0	5	1,01940	10	10,5
6	35,0	6	1,47483	5	29,8
7	101,0	2	1,92451	2	84,0
8	94,3	3	1,91985	3	83,1
9	27,8	7	1,37783	6	23,9
10	26,0	8	1,30138	8	20,0

[a]Trung bình của bốn lần lặp lại (không ghi nhận số liệu lặp lại gốc)

Ví dụ: Dùng số liệu phần trăm bông bạc (White Heads) từ trắc nghiệm 14 giống lúa, bố trí RCBD với ba lần lặp lại (Bảng 11.21). Khoảng biến động của số liệu từ 0 đến 26,39%. Do có nhiều giá trị nhỏ hơn 10, nên các số liệu được đổi sang $\sqrt{X+0,5}$ như ở Bảng 10.22. Phân tích phương sai được thực hiện dựa trên số liệu biến đổi của Bảng 10.22. Kết quả được ghi nhận trong Bảng 11.23.

Để so sánh tất cả các cặp trung bình, áp dụng kiểm định Duncan đối với các trung bình biến đổi và sau đó chuyển sang trung bình gốc. Phương pháp thực hiện tương tự như biến đổi logarith. Kết quả kiểm định Duncan được trình bày ở Bảng 11.24.

3. Đổi sang arcsin

Đổi sang arcsin thích hợp với các số liệu tỷ lệ, số liệu có được từ việc đếm và số liệu phần trăm. Chú ý, đây là số liệu phần trăm được tính từ số liệu đếm như phần trăm trái 1, 2, 3 hạt ở đậu nành,... Chớ không phải là dạng số liệu phần trăm không đến từ số liệu đếm như phần trăm protein, phần trăm carbohydrates.

Dùng bảng đổi arcsin (Phụ lục XIV) để đổi số liệu. Giá trị 0% nên thay bằng (1/4n) và giá trị 100% thay bằng (100 - 1/4n) với n là mẫu số dùng để tính phần trăm.

Không phải tất cả số liệu phần trăm đều đổi sang dạng arcsin. Ba nguyên tắc sau đây có thể được dùng để chọn dạng biến đổi số liệu phần trăm đến từ số liệu đếm.

* Nguyên tắc 1: Đối với số liệu phần trăm nằm trong khoảng biến động từ 30 đến 70%, không cần biến đổi.

* Nguyên tắc 2: Đối với số liệu phần trăm nằm trong khoảng biến động từ 0 đến 30% hoặc 70 đến 100%, nhưng không cả hai, nên đổi sang căn bậc hai.

* Nguyên tắc 3: Đối với số liệu phần trăm không theo khoảng biến động đã định ở nguyên tắc 1 hoặc nguyên tắc 2, nên chuyển sang arcsin.

Bảng 11.21: Phần trăm bông bạc (%) của 14 giống lúa được trắc nghiệm, bố trí RCBD với ba lần lặp lại. (Trích từ Bảng 7.19 của Gomez & Gomez. 1984. Trang 305)

Giống	Lặp lại I	Lặp lại II	Lặp lại III	Tổng (T)	T. bình
IR5	1,39	0,92	2,63	4,94	1,65
IR20-1	8,43	4,38	6,94	19,75	6,58
C4-C3-G	7,58	3,79	1,91	13,28	4,43
C168-134	8,95	12,81	3,22	24,98	8,33
BPI-76	4,16	17,39	8,06	29,61	9,87
MRC 407-1	4,68	1,32	2,09	8,09	2,70
PARC 2-2	2,37	5,32	4,86	12,55	4,18
TN1	0,95	0,70	0,98	2,63	0,88
Rexoro	26,09	25,36	15,69	67,14	22,38
Luma-1	26,39	22,29	1,98	50,66	16,89
IR127-80-1	21,99	12,88	5,15	40,02	13,34
IR1108-3-5	3,58	2,62	2,91	9,11	3,04
IR1561-228-3-3	0,19	0,00	0,61	0,80	0,27
IR2061-464-2	0,00	3,64	4,44	8,08	2,69
Tổng Rep.	116,75	113,42	61,47		
Tổng chung				291,64	

Bảng 11.22: Đổi số liệu ở Bảng 10.21 sang căn bậc hai, $\sqrt{X+0,5}$

Giống	Lặp lại I	Lặp lại II	Lặp lại III	Tổng (T)	Trung bình
IR5	1,37	1,19	1,77	4,33	1,44
IR20-1	2,99	2,21	2,73	7,93	2,64
C4-C3-G	2,84	2,07	1,55	6,46	2,15
C168-134	3,07	3,65	1,93	8,65	2,88
BPI-76	2,16	4,23	2,93	9,32	3,11
MRC 407-1	2,28	1,35	1,61	5,24	1,75
PARC 2-2	1,69	2,41	2,32	6,42	2,14
TN1	1,20	1,10	1,22	3,52	1,17
Rexoro	5,16	5,09	4,02	14,27	4,76
Luma-1	5,19	4,77	1,57	11,53	3,84
IR127-80-1	4,74	3,66	2,38	10,78	3,59
IR1108-3-5	2,02	1,77	1,85	5,64	1,88
IR1561-228-3-3	0,83	0,71	1,05	2,59	0,86
IR2061-464-2	0,71	2,03	2,22	4,96	1,65
Tổng Rep.	36,25	36,24	29,15		
Tổng chung				101,64	

Để giải thích dạng biến đổi sang arcsin, chúng ta dùng số liệu phần trăm rầy bông sống sót của thí nghiệm 12 giống lúa được bố trí theo thể thức RCBD với ba lần lặp lại (Bảng 11.25). Trên mỗi cây được đặt trong lồng, người ta thả vào 75 con và đếm số côn trùng còn sống sót. Vì số liệu phần trăm biến động trong khoảng từ 0 đến 100%, dựa trên nguyên tắc 3, dùng phương pháp đổi arcsin Trước khi biến đổi, tất cả giá trị 0 được thay bằng [1/4.(75)] và tất cả giá trị 100 bằng {100 - [1/4.(75)]}. Số liệu biến đổi

được trình bày trong Bảng 11.26. Phân tích phương sai như cách thông thường (Bảng 11.27) và kiểm định Duncan dựa trên trung bình biến đổi và sau đó chuyển sang trung bình gốc như phương pháp đổi sang logarith (Bảng 11.28).

Bảng 11.23: Phân tích phương sai của số liệu ở Bảng 11.22.

Nguồn biến động	Độ tự do	Tổng bình phương	Trung bình bình phương	F (tính)
Lặp lại	2	2,3971	1,1986	
Giống	13	48,0366	3,6951	5,88**
Sai số	26	16,3275	0,6280	
Tổng cộng	41	66,7612		

** = ý nghĩa ở 1%

Bảng 11.24: Áp dụng kiểm định Duncan để so sánh các trung bình nghiệm thức ở Bảng 11.21 qua so sánh Duncan của trung bình biến đổi ở Bảng 11.22

Giống	Trung bình biến đổi[a]	Trung bình gốc[a]
IR5	1,44 abc	1,65 abc
IR20-1	2,64 b-f	6,58 b- f
C4-C3-G	2,15 a-e	4,43 a-e
C168-134	2,88 c-f	8,33 c-f
BPI-76	3,11 def	9,87 def
MRC 407-1	1,75 a-d	2,70 a-d
PARC 2-2	2,14 a-e	4,18 a-e
TN1	1,17 ab	0,88 ab
Rexoro	4,75 g	22,38 g
Luma-1	3,84 fg	16,89 fg
IR127-80-1	3,59 efg	13,34 efg
IR1108-3-5	1,88 a- d	3,04 a-d
IR1561-228-3-3	0,86 a	0,27 a
IR2061-464-2	1,65 a-d	2,26 a-d

[a]Các trung bình có cùng chữ theo sau thì không khác biệt ý nghĩa ở mức 5%.

Bảng 11.25: Phần trăm sống sót (%) của rầy bông trên 12 giống lúa được trắc nghiệm theo kiểu bố trí hoàn toàn ngẫu nhiên (CRD) với ba lần lặp lại (Trích từ Bảng 7.23 của Gomez & Gomez. 1984. Trang 307).

Giống	Lặp lại I	Lặp lại II	Lặp lại III	Tổng (T)	Trung bình
ASD 7	44,00	25,33	48,00	117,33	39,11
Mudgo	21,33	49,33	80,00	150,66	50,22
Ptb 21	0,00	0,00	0,00	0,00	0,00
D 204-1	25,33	26,66	49,33	101,32	33,77
Su-Yai 20	24,00	26,66	54,66	105,32	35,11
Balamawee	0,00	0,00	20,00	20,00	6,67
DNJ 24	32,00	29,33	28,00	89,33	29,78
Ptb 27	0,00	0,00	0,00	0,00	0,00
Rathu Heenati	17,33	33,33	10,66	61,32	20,44
Taichung (N) 1	93,33	100,00	100,00	293,33	97,78
DS 1	13,33	36,00	33,33	82,66	27,55
BJ 1	46,66	46,66	16,00	109,32	36,44

Bảng 11.26: Đổi số liệu ở Bảng 11.25 sang Arcsin

Giống	Lặp lại I	II	III	Tổng (T)	Trung bình
ASD 7	41,55	30,22	43,85	115,62	38,54
Mudgo	27,51	44,62	63,43	135,56	45,19
Ptb 21	0,33	0,33	0,33	0,99	0,33
D 204-1	30,22	31,09	44,62	105,93	35,31
Su- Yai 20	29,33	31,09	47,67	108,09	36,03
Balamawee	0,33	0,33	26,57	27,23	9,08
DNJ 24	34,45	32,79	31,95	99,19	33,06
Ptb 27	0,33	0,33	0,33	0,99	0,33
Rathu Heenati	24,60	35,26	19,06	78,92	26,31
taichung (N) 1	75,03	89,67	89,67	254,37	84,79
DS 1	21,41	36,87	35,26	93,54	31,18
BJ 1	43,08	43,08	23,58	109,74	36,58

Bảng 11.27: Phân tích phương sai của số liệu ở Bảng 11.26

Nguồn biến động	Độ tự do	Tổng bình phương	Trung bình bình phương	F (tính)
Giống	11	16838,6368	1530,7852	16,50**
Sai số	24	2225,9723	92,7488	
Tổng cộng	35	19064,6091		

Bảng 11.28: Áp dụng kiểm định Duncan để so sánh các trung bình nghiệm thức ở Bảng 11.25 qua so sánh Duncan của trung bình biến đổi ở Bảng 11.26

Giống	Trung bình biến đổi	Trung bình gốc
ASD 7	38,54 bc	39,11 bc
Mudgo	45,19 c	50,22 c
Ptb 21	0,33 a	0,00 a
D 204-1	35,31 bc	33,77 bc
Su- Yai 20	36,03 bc	35,11 bc
Balamawee	9,08 a	6,67 a
DNJ 24	33,06 bc	29,78 bc
Ptb 27	0,33 a	0,00 a
Rathu Heenati	26,31 b	20,44 b
taichung (N) 1	84,79 d	97,78 d
DS 1	31,18 bc	27,55 bc
BJ 1	36,58 bc	36,44 bc

[a]Các trung bình có cùng chữ theo sau thì không khác biệt ý nghĩa ở mức 5%.

CHƯƠNG 12

TRÌNH BÀY CÁC KẾT QUẢ NGHIÊN CỨU

Công việc sau cùng và quan trọng nhất đối với bất cứ nhà nghiên cứu nào là tóm tắt và trình bày các kết quả nghiên cứu. Mục đích chính của công việc này là tìm cách trình bày tất cả các phần quan tâm dưới dạng có thể dễ hiểu. Số liệu thô không được tóm tắt vẫn còn tối nghĩa và không có giá trị trong các file nghiên cứu. Số liệu được trình bày đúng sẽ tạo cơ hội tốt nhất cho người sử dụng nghiên cứu sâu.

Trình bày đúng các kết quả nghiên cứu là một công việc được nhiều nhà nghiên cứu quan tâm, song cách trình bày như thế nào được xem là thích hợp thì chưa có tài liệu hướng dẫn. Mặc dù trong hầu hết các lĩnh vực khoa học, có nhiều mô tả khổ và kiểu mẫu bài báo cáo đã được công bố, nhưng các hướng dẫn đó chưa được tiêu chuẩn hóa và đồng bộ. Mỗi tác giả thường chọn cho mình phương pháp tốt nhất để trình bày số liệu. Một số dạng bảng biểu và đồ biểu của các kết quả nghiên cứu được trình bày dưới đây được tham khảo theo hướng dẫn của Gomez and Gomez (1984).

Nhìn chung, có hai dạng thông tin được trình bày trong nghiên cứu nông nghiệp:

- Dạng thông tin thứ nhất là mô tả các vật liệu thí nghiệm và các yếu tố môi trường đối với địa điểm thí nghiệm. Trình bày đúng bao gồm việc chọn các đặc tính để trình bày, dạng thống kê sử dụng và số đo chính xác thích hợp. Đối với các đặc tính riêng biệt, số trung bình số học là số thống kê được sử dụng phổ biến nhất, kèm theo sai số chuẩn của trung bình như chỉ số chính xác (Bảng 12.1). Trong khi, đối với các đặc tính đo đạc qua thời gian dài như lượng mưa và bức xạ mặt trời, đồ biểu hình thanh và đồ thị đường (thẳng hoặc gảy khúc) hoặc cả hai được dùng phổ biến (Hình 12.1).

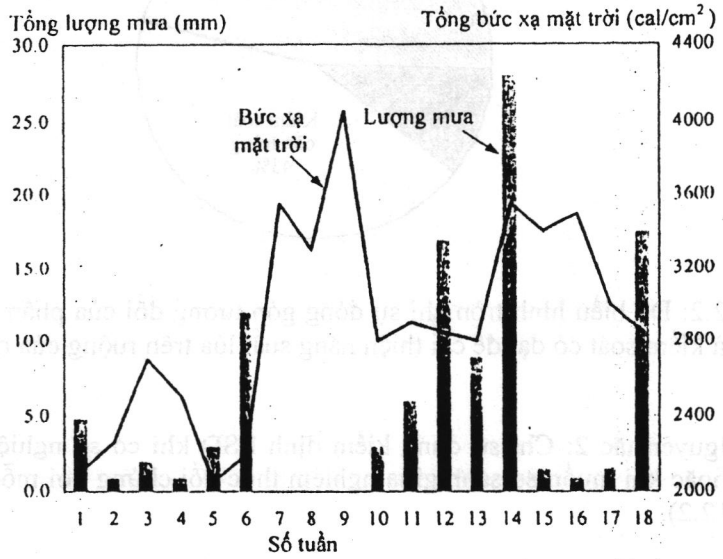
Bảng 12.1: Tính chất hóa-lý của lớp đất mặt ở địa điểm thí nghiệm phân bón dài hạn

Đặc tính đất	Đơn vị	Trung bình
PH	-	5,8 ± 0,1
Chất hữu cơ	g	4,15 ± 0,01
Đạm tổng số	g	0,31 ± 0,01
Lân dễ tiêu	ppm	7,3 ± 1,0
K trao đổi	meq/100g	1,46 ± 0,07
Calcium trao đổi	meq/100g	9,18 ± 0,27
Khả năng trao đổi cation (EC)	meq/100g	73,3 ± 0,6

- Dạng thông tin thứ hai là trình bày kết quả phân tích số liệu. Cách trình bày thay đổi nhiều theo nghiệm thức được trắc nghiệm (nghĩa là, nghiệm thức định tính hoặc định lượng và nghiệm thức của thí nghiệm một nhân tố hay thí nghiệm thừa số), các tính trạng được đo lường (nghĩa là, các số đo qua thời gian dài hoặc số liệu nhiều tính trạng), và phương pháp thống kê được sử dụng (nghĩa là, so sánh trung bình, hồi qui hoặc kiểm định chi bình phương).

Dạng bảng là dạng được sử dụng phổ biến nhất để trình bày các kết quả nghiên cứu. Bảng thì mềm dẽo và có thể được dùng để trình bày tính đa dạng các kết quả. Thực tế, bất cứ dạng số liệu nào không thể tóm tắt bằng đồ biểu đều có thể trình bày dạng bảng. So sánh trung bình của các nghiệm thức riêng biệt (Bảng kiểm định LSD,

Duncan,...), và mô tả một vài tính chất mấu chốt của môi trường (Bảng 12.1) là hai cách sử dụng dạng bảng phổ biến.



Hình 12.1: Lượng mưa và bức xạ mặt trời trong suốt giai đoạn thí nghiệm bằng cách phối hợp đồ biểu hình thanh và đồ thị đường.

Trình bày bằng đồ biểu được sử dụng phổ biến nhất trong nghiên cứu nông nghiệp, được xếp hạng theo mức độ chính xác và tần số sử dụng, là đồ thị đường, đồ biểu hình thanh và hình tròn. Đồ thị đường thích hợp để trình bày mối quan hệ giữa hai biến định lượng (biến liên tục) như quan hệ giữa sự đáp ứng của cây trồng và biến nghiệm thức định lượng. Đồ biểu hình thanh nhìn chung được sử dụng cho số liệu rời (không liên tục) như phân bố tần số và số liệu phần trăm. Đồ biểu hình tròn được dùng để trình bày sự khác nhau nổi bật về tầm quan trọng tương đối của một vài thành phần trong toàn thể đơn vị (Hình 12.2).

I. THÍ NGHIỆM MỘT NHÂN TỐ

Trong thí nghiệm một nhân tố có hai dạng so sánh giữa các trung bình nghiệm thức: so sánh cặp cho các nghiệm thức riêng biệt và so sánh hướng cho các nghiệm thức định lượng.

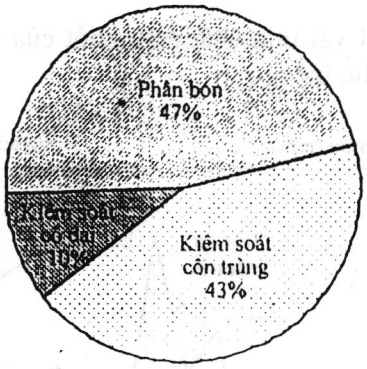
1. Các nghiệm thức riêng biệt

Kiểm định LSD, Duncan và Newman Keuls là các phương pháp được sử dụng phổ biến nhất để so sánh các trung bình của nghiệm thức riêng biệt. Các kết quả so sánh như thế có thể được trình bày hoặc dạng bảng hoặc đồ biểu hình thanh.

a. Dạng bảng với kiểm định LSD

Trình bày so sánh trung bình dựa trên kiểm định LSD thì đơn giản, và nên sử dụng bất cứ lúc nào kiểm định LSD có giá trị. Một vài nguyên tắc để trình bày đúng so sánh trung bình bằng cách sử dụng kiểm định LSD là:

- Nguyên tắc 1: Chỉ sử dụng kiểm định LSD khi kiểm định F trong phân tích phương sai có ý nghĩa.



Hình 12.2: Đồ biểu hình tròn chỉ sự đóng góp tương đối của phân bón, kiểm soát côn trùng và kiểm soát cỏ dại để cải thiện năng suất lúa trên ruộng của nông dân.

- Nguyên tắc 2: Chỉ sử dụng kiểm định LSD khi có số nghiệm thức từ năm trở xuống hoặc khi muốn so sánh giữa nghiệm thức đối chứng với mỗi nghiệm thức khác (Bảng 12.2).

Bảng 12.2: Trình bày bảng để giải thích cách sử dụng kiểm định LSD trong việc so sánh năng suất của ba giống bắp lai triển vọng A, B, và D với giống đối chứng C.

Giống bắp	Năng suất trung bình (t/ha)*
A	1,46
B	1,47
C (Đối chứng)	1,07
D	1,34
$LSD_{0,05}$	0,25

*Trung bình của 4 lần lặp lại

- Nguyên tắc 3: Chỉ sử dụng một dạng kiểm định. Không dùng cả hai kiểm định LSD và Duncan cho cùng một bộ trung bình nghiệm thức.

- Nguyên tắc 4: Giá trị LSD phải có cùng đơn vị đo lường và cùng số lẻ như các giá trị trung bình.

- Nguyên tắc 5: Khi thực hiện phân tích phương sai với số liệu đã biến đổi, kiểm định LSD chỉ có thể được trình bày nếu các giá trị trung bình được trình bày ở tỷ lệ biến đổi.

- Nguyên tắc 6: Để so sánh tất cả các cặp trung bình có thể có, trình bày giá trị LSD hoặc được ghi ở hàng cuối của bảng hoặc chú thích cuối trang (Bảng 12.2). Để so sánh giữa nghiệm thức đối chứng với mỗi nghiệm thức khác, đánh dấu mỗi nghiệm thức với **, * hoặc ns (hoặc không đánh dấu) tùy thuộc vào mức ý nghĩa của kiểm định LSD (xem Chương 8, Bảng 8.1).

b. Dạng bảng với kiểm định Duncan

Một vài nguyên tắc để sử dụng đúng và trình bày kiểm định Duncan:

- Nguyên tắc 1: Để so sánh tất cả cặp nghiệm thức có thể có, sử dụng kiểm định Duncan khi số nghiệm thức trên năm. Trường hợp có ít hơn sáu nghiệm thức, có thể sử dụng kiểm định LSD. Khi sử dụng số liệu biến đổi trong phân tích phương sai, và các

trung bình nghiệm thức được trình bày ở tỷ lệ gốc, sử dụng kiểm định Duncan bất chấp số nghiệm thức.

- Nguyên tắc 2: Dùng ký hiệu alphabet

* Bắt đầu chữ đầu tiên cho mỗi trung bình ở cùng vị trí.

* Nếu hơn hai nghiệm thức có từ bốn chữ trở lên, dùng ký hiệu gạch nối để rút ngắn chuỗi các chữ. Ví dụ, chuỗi abcd được viết là a-d, chuỗi bcdefg được viết là b-g.... (Bảng 12.3).

Bảng 12.3: So sánh Duncan cho các trung bình biến đổi. Trình bày trung bình ở tỷ lệ gốc.

Giống	Trung bình biến đổi	Trung bình gốc
IR5	1,44 abc	1,65 abc
IR30-1	2,64 b-f	6,58 b-f
C4-63G	2,15 a-c	4,43 a-c
C168-134	2,88 c-f	8,33 c-f
BPI-26	3,11 def	9,87 def
MRC407-1	1,75 a-d	2,70 a-d
PARC2-2	2,14 a-c	4,18 a-c
TNT	1,17 ab	0,88 ab
Resono	4,26 g	22,38 g
Luma-1	3,84 fg	16,89 fg
IR127-80-1	3,59 efg	13,34 efg
IR1108-3-5	1,88 a-d	3,04 a-d
IR1561-228-3-3	0,86 a	0,27 a
IR3061-464-2	1,65 a-d	2,69 a-d

- Nguyên tắc 3: Khi chỉ có một phần nghiệm thức được trình bày trong bảng, cần điều chỉnh chữ kiểm định Duncan để các chữ được liên tục và sử dụng ít chữ (Bảng 12.4).

Bảng 12.4: Điều chỉnh chữ Duncan khi chỉ trình bày các trung bình của sáu trong tổng số 10 giống được trắc nghiệm.

	Năng suất				Chiều cao cây			
	T.B.	DUNCAN			T.B.	DUNCAN		
Giống	t/ha	Gốc	CĐC	ĐĐC	cm	Gốc	CĐC	ĐĐC
CPM-13-32-41	1,16	a	a	a	56,5	f	f	d
CR214-75-52-102	0,44	c	c	b	53,2	f	f	d
D6-2-2	0,98	ab			120,2	a		
Bala	1,16	a			72,6	d		
Amapuma	0,49	c	c	b	64,7	e	e	c
CR245-1	1,14	a	a	a	105,9	b	b	a
OR165-28-14-B	0,60	bc	bc	b	66,3	de	de	c
IAC	0,62	bc			104,6	b		
KS109	0,51	c			67,4	de		
IR12787-3	0,81	abc	abc	ab	84,3	c	c	b

Trung bình của 4 lần lặp lại. Sử dụng kiểm định DUNCAN ở mức ý nghĩa 5%.
Các giống in đậm được loại khỏi bảng.
CĐC = Chưa điều chỉnh và ĐĐC = Đã điều chỉnh.

- Nguyên tắc 5: Các trung bình có cùng giá trị không nên trình bày các chữ Duncan khác nhau. Các trường hợp như thế thường xảy ra do việc làm tròn số dẫn đến số lẻ bị giảm. Khi điều đó xảy ra, nên gia tăng số lẻ của các giá trị trung bình (Bảng 12.5). Chú ý, có thể trình bày các giá trị trung bình có số lẻ nhiều hơn số liệu gốc.

Bảng 12.5: Điều chỉnh số lẻ để tránh xảy ra trường hợp các chữ khác nhau lại có cùng giá trị trung bình.

Số	Số cành hữu hiệu*	
Nghiệm thức	Trước khi điều chỉnh	Sau khi điều chỉnh
1	1 ab	1,1 ab
2	1 ab	1,2 ab
3	1 ab	1,3 ab
4	1 a	1,4 a
5	1 ab	1,2 ab
6	1 b	0,6 b
7	2 a	1,5 a

*Trung bình của 4 lần lặp lại. Sử dụng kiểm định DUNCAN ở mức ý nghĩa 5%.

- Nguyên tắc 6: Giảm bớt số lẻ nếu các chữ trong kiểm định Duncan khác nhau. Ví dụ, trong Bảng 12.6 không cần thiết dùng số lẻ ở các giá trị trung bình vì có sự khác biệt ý nghĩa giữa các trung bình (sự khác biệt này không thể phát hiện dưới 10%).

Bảng 12.6: Điều chỉnh số lẻ để phản ánh độ chính xác được chứng tỏ qua kiểm định Duncan.

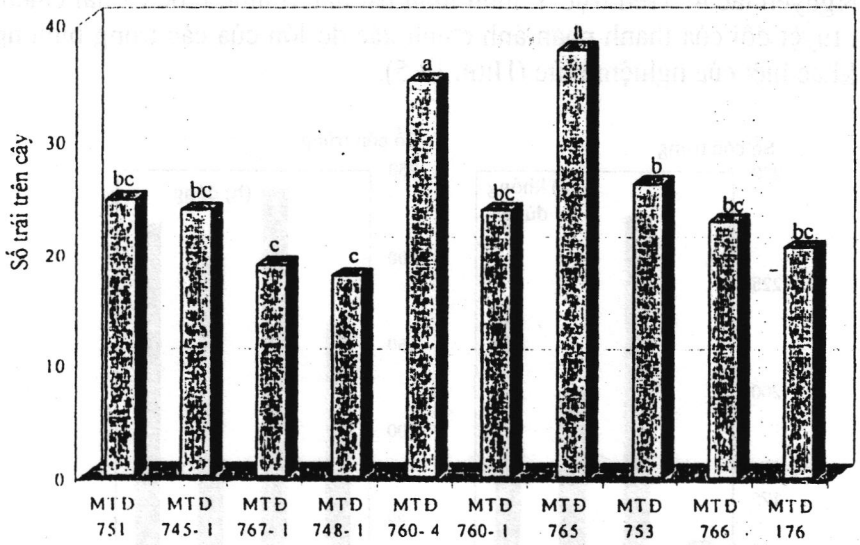
Số	Số cành hữu hiệu*	
Nghiệm thức	Trước khi điều chỉnh	Sau khi điều chỉnh
1	17.4 b	17 b
2	23,5 b	24 b
3	21,3 b	21 b
4	51,6 a	52 a
5	54,7 a	55 a
6	55,7 a	56 a
7	57,5 a	58 a
8	58,2 a	58 a
9	61,4 a	61 a

*Trung bình của 4 lần lặp lại. Sử dụng kiểm định Duncan ở mức ý nghĩa 5%.

- Nguyên tắc 7: Không trình bày kiểm định Duncan khi tất cả các trung bình có cùng chữ. Chú thích cuối trang chứng tỏ không có sự khác biệt ý nghĩa giữa các trung bình nghiệm thức là đủ.

c. Đồ biểu hình thanh

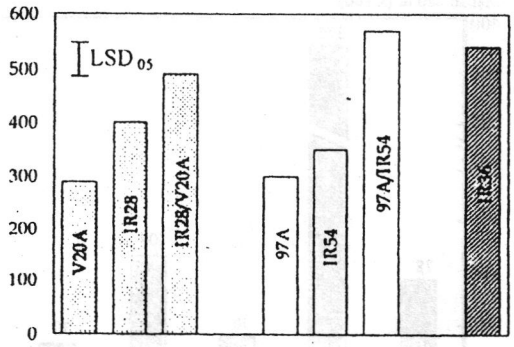
Đồ biểu hình thanh thích hợp cho các nghiệm thức riêng biệt và tương đối ít. Nhìn chung, đồ biểu hình thanh được dùng để nhấn mạnh sự khác biệt nổi bật giữa các nghiệm thức (Hình 12.3) hoặc để chỉ một vài kiểu thay đổi đặc biệt giữa các nhóm nghiệm thức (Hình 12.4).



Hình 12.3: Đồ biểu hình thanh chỉ sự so sánh số trái trên cây của 10 giống đậu nành. Các thanh có chữ Duncan khác nhau khác biệt ở mức ý nghĩa 5%.

Một vài nguyên tắc sử dụng đồ biểu hình thanh để trình bày các kết quả so sánh trung bình bao gồm các nghiệm thức riêng biệt là:

- Nguyên tắc 1: Sử dụng đồ biểu hình thanh khi cần nhấn mạnh sự khác biệt nổi bật hoặc kiểu thay đổi tương đối, và khi không cần thiết để duy trì mức độ chính xác cao của từng giá trị trung bình.

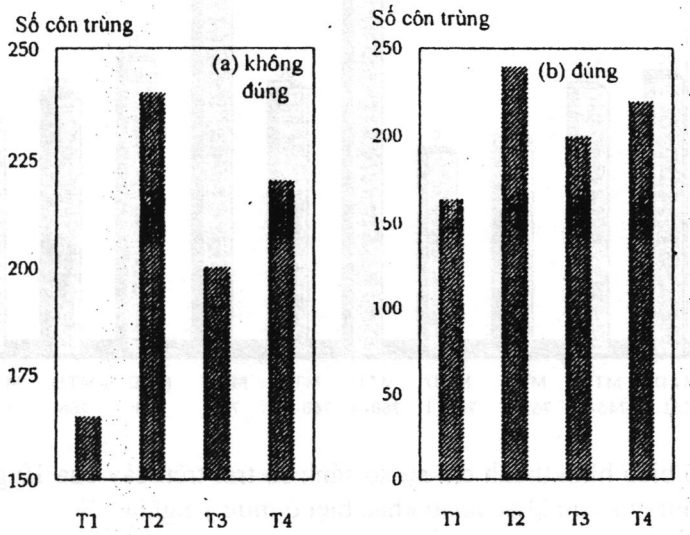


Hình 12.4: Đồ biểu hình thanh chỉ năng suất tương đối của các giống lúa lai, cha mẹ của chúng và giống đang trồng phổ biến IR36. Các thanh được gom nhóm và trình bày giá trị LSD.

- Nguyên tắc 2: Khi áp dụng kiểm định Duncan, sử dụng ký hiệu alphabet bằng cách đặt các chữ trên đỉnh của mỗi thanh nghiệm thức (Hình 12.3).

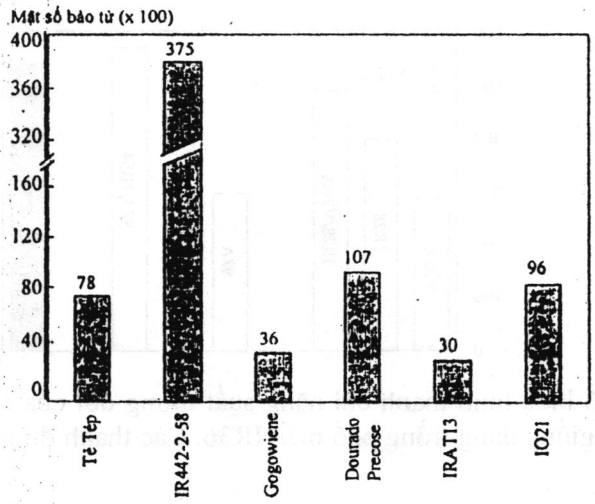
- Nguyên tắc 3: Khi dùng kiểm định LSD. Biểu diễn độ lớn của LSD như được giải thích ở Hình 12.4.

- Nguyên tắc 4: Trên trục Y luôn luôn bắt đầu ở mức 0 để cả hai chiều cao tương đối và tuyệt đối của thanh phản ánh chính xác độ lớn của các trung bình nghiệm thức và sự khác biệt của nghiệm thức (Hình 12.5).



Hình 12.5: Đồ biểu hình thanh giải thích phương pháp đúng (b) trục Y bắt đầu ở mức 0. Phương pháp không đúng (a) rút ngắn trục Y và phóng đại sự khác biệt giữa các nghiệm thức.

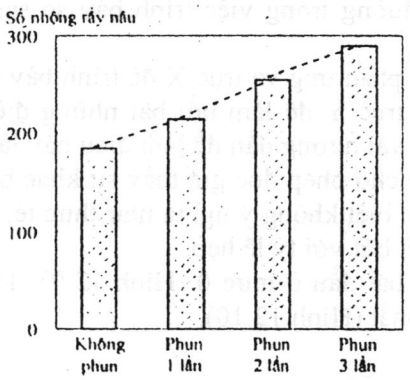
- Nguyên tắc 5: Tránh sử dụng thanh ngắt đoạn nhằm mục đích rút ngắn chiều cao khi có sự khác biệt cực lớn ở chiều cao của thanh (Hình 12.6).



Hình 12.6: Đồ biểu hình thanh chỉ việc sử dụng thanh bị cắt ngang để rút ngắn chiều cao và đặt số liệu trên đỉnh của mỗi thanh. Cả hai đều nên tránh.

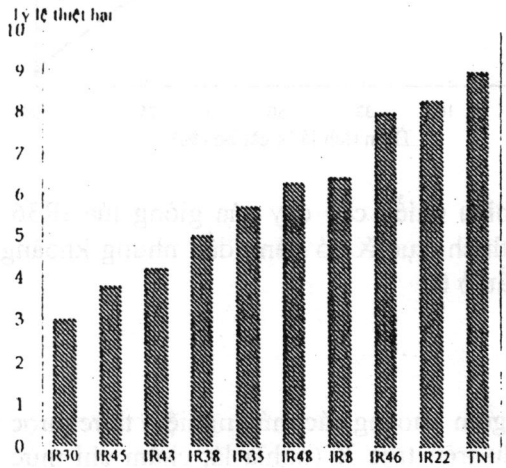
- Nguyên tắc 6: Tránh viết số liệu thực trên đỉnh của mỗi thanh (Hình 12.6). Ở đồ biểu hình thanh, giá trị trung bình nên được đọc từ trục Y dựa trên chiều cao tương ứng của thanh. Nếu đòi hỏi mức độ chính xác cao hơn có thể đạt được từ đồ biểu hình thanh thì đồ biểu hình thanh không phải là dạng trình bày thích hợp.

- Nguyên tắc 7: Tránh nối đỉnh của các thanh kề nhau bằng một đường (Hình 12.7). Đường nối các thanh như thế để chỉ hướng thì không thích hợp đối với biến không liên tục (biến rời). Nếu biến trên trục X là biến liên tục, nên sử dụng đồ thị đường và không dùng đồ biểu hình thanh.



Hình 12.7: Giải thích sự phối hợp đồ biểu hình thanh và đồ thị đường thẳng đối một tính trạng duy nhất. Điều này nên tránh.

- Nguyên tắc 8: Xác định chuỗi liên tục các thanh dựa trên dạng nghiệm thức và điểm đặc biệt cần được nhấn mạnh. Khi có sự gom nhóm tự nhiên các nghiệm thức, các thanh có thể được trình bày theo các nhóm nghiệm thức. Mặt khác, các thanh có thể được sắp xếp theo thứ hạng của chúng - từ cao nhất đến thấp nhất hoặc ngược lại (Hình 12.8).



Hình 12.8: Đồ biểu hình thanh chỉ tỷ lệ thiệt hại tương đối của rầy xanh trên các giống lúa. Biểu đồ chỉ sự sắp xếp các thanh theo thứ hạng.

2. Nghiệm thức định lượng: Đồ thị đường

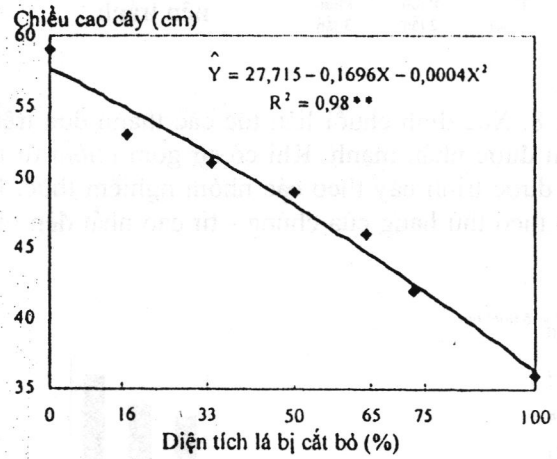
So sánh hướng, chỉ mối quan hệ hàm số giữa mức nghiệm thức và phản ứng sinh học tương ứng, là phương pháp so sánh trung bình thích hợp nhất giữa các nghiệm thức định lượng. Đối với các so sánh như thế, đồ thị đường sẽ thích hợp nhất vì các giá trị của phản ứng được quan tâm không bị giới hạn đối với các mức nghiệm thức được kiểm định nhưng cũng trên tất cả các điểm trong biên độ các nghiệm thức được trắc nghiệm. Chú ý với các bảng hoặc đồ biểu hình thanh, phản ứng chỉ định rõ cho mỗi mức nghiệm thức được trắc nghiệm thật sự. Nhưng với đồ thị đường, phản ứng được định rõ cho tất cả các điểm giữa các mức nghiệm thức cao nhất và thấp nhất.

Một vài nguyên tắc để sử dụng đồ thị đường trong việc trình bày so sánh trung bình các nghiệm thức định lượng là:

- Nguyên tắc 1: Dùng trục Y để trình bày phản ứng và trục X để trình bày các mức nghiệm thức. Chọn các tỷ lệ cho trục X và trục Y để làm nổi bật những điểm quan trọng, nhưng tránh làm méo mó kết quả. Một vài hướng dẫn để làm điều này là:

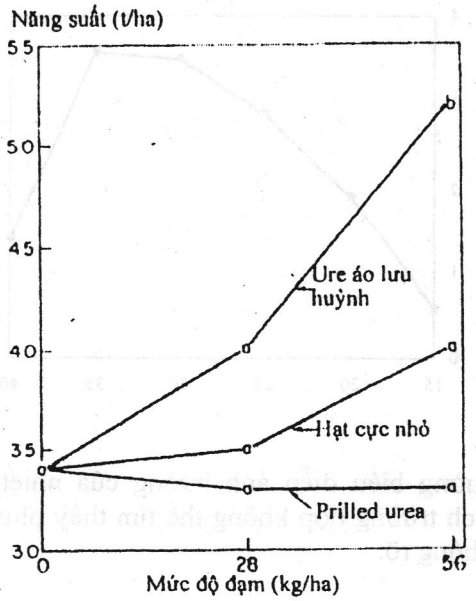
* Chọn tỷ lệ trên trục Y như thế nào để cho phép đọc giả thấy sự khác biệt được chứng tỏ là có ý nghĩa và không thấy sự khác biệt không ý nghĩa như thực tế. Sự khác biệt được phóng đại với tỷ lệ rộng và được rút bớt với tỷ lệ hẹp.

* Với đồ thị đường, trục Y không cần bắt đầu ở mức 0 (Hình 12.9). Tương tự, trục X nên bắt đầu ở mức nghiệm thức thấp nhất (Hình 12.10).



Hình 12.9: Đồ thị đường biểu chiều cao cây của giống lúa IR36 bị ảnh hưởng bởi phạm vi lá bị cắt bỏ: giải thích trục X có đánh dấu nhưng khoảng cách không bằng nhau và trục Y không bắt đầu ở 0.

* Trục X chỉ nên bao gồm khoảng các mức nghiệm thức được trắc nghiệm. Càng nhiều càng tốt, sự đánh dấu trên trục X (nghĩa là, chấm chỉ mức nghiệm thức) nên tương ứng với mức nghiệm thức (Hình 12.10). Tuy nhiên, nếu số nghiệm thức lớn, không cần đánh dấu tất cả các mức nghiệm thức (Hình 12.12), hoặc nếu các nghiệm thức cách khoảng không đều nhau, không cần đánh dấu cách đều (Hình 12.10). Các dấu trên trục Y nên cách đều. Giảm tối thiểu số dấu để các dấu cần thiết cho khả năng đọc rõ ràng.



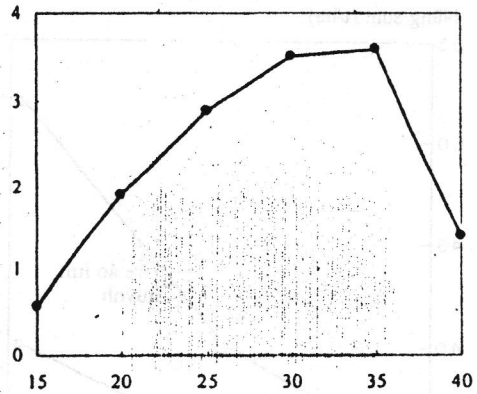
Hình 12.10: Đồ thị đường biểu diễn năng suất trung bình của giống lúa RD19 chịu tác động bởi mức độ và dạng ure. Giải thích trường hợp không thể rút ra phương trình hồi qui do các mức nghiệm thức không bằng nhau và sử dụng kiểm định Duncan. Các trung bình có cùng chữ khác biệt không ý nghĩa ở mức 5%.

- Nguyên tắc 2: Dùng đồ thị đường khi có ít nhất ba nghiệm thức.
- Nguyên tắc 3: Bất cứ lúc nào có thể, ước lượng phương trình hồi qui thích hợp phải chắc chắn đồ thị đường bao gồm:
  * Các điểm quan sát (nghĩa là, số liệu sử dụng phù hợp đường hồi qui)
  * Đường hồi qui ước lượng, hoặc đường cong, được vẽ trên biên độ các mức nghiệm thức được trắc nghiệm.
  * Hàm hồi qui ước lượng, mức ý nghĩa và hệ số tương quan.
- Nguyên tắc 4: Khi không thể rút ra một phương trình hồi qui chính xác, hoặc do số nghiệm thức không đủ (Hình 12.10) hoặc do mối quan hệ hàm số không ý nghĩa (Hình 12.11), đơn giản vẽ một đường để nối các điểm quan sát và kết hợp kiểm định ý nghĩa thích hợp.

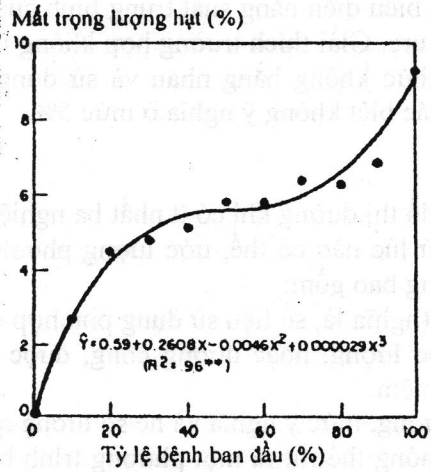
II. THÍ NGHIỆM THỪA SỐ

Đối với thí nghiệm thừa số, trình bày so sánh trung bình nên dựa trên các điểm sau đây:

- Tương tác giữa các nhân tố trắc nghiệm. Khi tương tác giữa các nhân tố trắc nghiệm có ý nghĩa và lớn, nên trình bày nhấn mạnh bản chất và độ lớn của tương tác. Điều này có thể được thực hiện bằng cách lập một đồ biểu nhiều chiều hoặc một bảng nhiều chiều bao gồm các nhân tố tương tác. Ví dụ, nếu hai nhân tố riêng biệt, A và B tương tác, lúc đó nên trình bày bảng hai chiều bao gồm nhân tố A và B.



Hình 12.11: Đồ thị đường biểu diễn ảnh hưởng của nhiệt độ trên sự oxy hoá α-naphthylamine: giải thích trường hợp không thể tìm thấy phương trình hồi qui do bản chất của mối quan hệ không rõ.



Hình 12.12: Đồ thị đường biểu diễn sự mất trọng lượng được ước lượng do bệnh: giải thích trục X có dấu ít hơn các mức bệnh được trắc nghiệm.

- Dạng nhân tố kiểm định. Như ở trường hợp của thí nghiệm một nhân tố, cách trình bày số liệu thích hợp là một bảng hoặc đồ biểu hình thanh cho các nhân tố riêng biệt và một đồ thị đường khi ít nhất một trong những nhân tố là định lượng.
- Số nhân tố trắc nghiệm. Trình bày số liệu gia tăng tính phức tạp khi số nhân tố trắc nghiệm gia tăng. Khi số nhân tố trắc nghiệm (k) không nhiều hơn ba, nhìn chung có thể sử dụng một bảng hoặc đồ biểu k chiều. Mặt khác, chỉ có những nhân tố có tương tác lẫn nhau mới được chứa trong cùng một bảng hoặc đồ biểu.

1. Dạng bảng

Nguyên tắc lập một bảng nhiều chiều để trình bày số liệu của một thí nghiệm thừa số là:

- Nguyên tắc 1: Dùng dạng bảng khi tất cả các nhân tố được trình bày là riêng biệt. Ngược lại, khảo sát bằng cách sử dụng đồ thị đường.

- Nguyên tắc 2: Tránh trình bày số liệu của thí nghiệm thừa số trong bảng một chiều (Bảng 12.7). Dạng bảng như thế không dễ đánh giá tương tác giữa các nhân tố trắc nghiệm.

Bảng 12.7: Trình bày không đúng số liệu của thí nghiệm thừa số trong bảng một chiều

Nghiệm thức			Năng suất
Vôi	Giống	MnO_2	trung bình (t/ha)*
Không	IR26	Không	3,6 d
		Có	3,9 d
	IR43	Không	4,0 cd
		Có	6,2 a
Có	IR26	Không	4,3 cd
		Có	4,8 bcd
	IR43	Không	5,3 b
		Có	6,2 a

*Trung bình của 4 lần lặp lại. Sử dụng kiểm định Duncan ở mức ý nghĩa 5%

- Nguyên tắc 3: Số nhân tố được chứa trong một bảng nhiều chiều nên được xác định như sau:

* Bất cứ lúc nào có thể, chiều của bảng nên bằng số nhân tố trắc nghiệm. Điều này thường được thực hiện khi số nhân tố không nhiều hơn ba và các mức độ của mỗi nhân tố không quá lớn (Bảng 12.8).

Bảng 12.8: Bảng ba chiều của các trung bình để trình bày số liệu của thí nghiệm ba nhân tố (Dạng Bảng này thích hợp hơn Bảng 12.7)

	Năng suất trung bình (t/ha)[a]			
	IR26		IR43	
MnO_2	Có vôi	Không vôi	Có vôi	Không vôi
Có	4,8 bcd	3,9 d	6,2 a	6,2 a
Không	4,3 cd	3,6 d	5,3 b	4,0 cd

[a]Trung bình của 4 lần lặp lại. Sử dụng kiểm định Duncan ở mức ý nghĩa 5%

* Khi có nhiều hơn một tương tác của cùng bậc có ý nghĩa, lập một bảng cho mỗi tương tác hoặc một bảng bao gồm tất cả các nhân tố có tương tác. Ví dụ, nếu tương tác A x B và A x C có ý nghĩa, hoặc lập hai bảng trung bình hai chiều, A x B và A x C, hoặc trình bày bảng trung bình ba chiều A x B x C.

- Nguyên tắc 4: Bất cứ lúc nào một hay nhiều nhân tố chứa trong cùng một bảng chỉ có hai mức độ cho mỗi nhân tố, trình bày sự khác biệt giữa hai mức độ thay vì các trung bình (Bảng 12.9). Cách này dễ đánh giá độ lớn cũng như mức ý nghĩa của ảnh hưởng của mỗi nhân tố. Ví dụ, từ Bảng 12.9, có thể dễ dàng đi đến các kết luận sau đây:

* Ở giống IR28, ảnh hưởng của vôi hoặc MnO_2 không ý nghĩa.

* Ở giống IR43, ảnh hưởng của MnO_2 gia tăng khi không bón vôi và ảnh hưởng của vôi chỉ quan sát được khi không dùng MnO_2.

Bảng 12.9: Bảng trung bình để trình bày số liệu của thí nghiệm thừa số 2^3. Chỉ sự khác biệt giữa hai mức độ của mỗi nhân tố.

	Năng suất trung bình (t/ha)[a]					
	IR26			IR43		
MnO_2	Có vôi	Không vôi	Khác biệt	Có vôi	Không vôi	Khác biệt
Có	4,8	3,9	$0,9^{ns}$	6,2	6,2	0,0
Không	4,3	3,6	$0,7^{ns}$	5,3	4,0	1,3*
Khác biệt	$0,5^{ns}$	$0,3^{ns}$		0,9*	2,2**	

[a]Trung bình của 4 lần lặp lại. ** = Khác biệt ở mức ý nghĩa 1%, * Khác biệt ở mức ý nghĩa 5%, ^{ns} = Khác biệt không ý nghĩa.

- Nguyên tắc 5: Với bố trí khối hoàn toàn (CRD, RCBD hoặc hình vuông latin), dùng ký hiệu chữ để trình bày kết quả kiểm định Duncan khi thực hiện so sánh trung bình, hoặc theo hàng hoặc theo cột, cho tất cả các nghiệm thức (Bảng 12.8).

- Nguyên tắc 6: Đối với thí nghiệm hai nhân tố theo kiểu bố trí lô phụ *(Split plot)* hoặc dải phụ *(Strip plot)*, tiêu chuẩn kiểm định cho nhân tố hàng khác với tiêu chuẩn kiểm định cho nhân tố cột. Các hướng dẫn để trình bày kết quả của các dạng số liệu như thế là:

* Nếu tương tác A x B có ý nghĩa và số mức độ của nhân tố A ít hơn sáu trong khi số mức độ của nhân tố B thì không, ấn định nhân tố B như nhân tố cột và nhân tố A như nhân tố hàng (Bảng 12.10). Đặt chữ Duncan thích hợp để so sánh các trung bình của nhân tố A ở mỗi mức độ của nhân tố B. Để so sánh các trung bình của nhân tố B ở mỗi mức độ của nhân tố A, dùng kiểm định LSD và trình bày giá trị LSD thích hợp như một chú thích ở cuối bảng (Bảng 12.10).

* Nếu tương tác A x B có ý nghĩa và số mức độ của cả hai nhân tố A và B đều ít hơn sáu, nhưng một nhân tố, chẳng hạn A, thì quan trọng hơn hoặc cần nhấn mạnh nhiều hơn nhân tố kia, dùng kiểm định Duncan trên nhân tố A và kiểm định LSD trên nhân tố B (Bảng 12.11). Nếu cả hai nhân tố cần được nhấn mạnh như nhau, dùng kiểm định LSD cho cả hai nhân tố và qui định các giá trị LSD thích hợp, một cho mỗi nhân tố, chú thích ở cuối bảng (Bảng 12.12).

* Nếu tương tác A x B có ý nghĩa và cả hai nhân tố A và B có hơn năm mức độ cho mỗi nhân tố, nhưng một nhân tố, dùng hai bộ chữ Duncan, một cho nhân tố cột và một cho nhân tố hàng (Bảng 12.13).

* Nếu tương tác A x B có ý nghĩa và một trong hai nhân tố, chẳng hạn A, có hai mức độ và nhân tố khác, chẳng hạn B, có từ sáu mức độ trở lên, ấn định nhân tố A như nhân tố cột và trình bày sự khác biệt giữa hai mức độ của A theo nguyên tắc 4. Đối với nhân tố B, dùng kiểm định Duncan (Bảng 12.14).

* Nếu tương tác A x B không ý nghĩa và trình bày bảng trung bình A x B, đơn giản, dùng kiểm định Duncan để so sánh các trung bình A được tính trung bình trên tất cả các mức độ của nhân tố B và các trung bình B được tính trung bình trên tất cả các mức độ của nhân tố A (Bảng 12.15). Chú ý, trong trường hợp này, cùng bộ chữ Duncan (a, b, c,...) có thể được dùng cho cả hai nhân tố.

Bảng 12.10: Ảnh hưởng của việc kiểm soát cỏ dại và phương pháp chuẩn bị đất trên năng suất đậu xanh. Giải thích cách sử dụng kiểm định Duncan cho nhân tố cột và kiểm định LSD cho nhân tố hàng.

	Năng suất trung bình (kg/ha)[a]		
		Luống hạt cũ với	
Kiểm soát cỏ dại	Theo lối cổ truyền	Glyphosate	Một lần cày xoay
Trifluradin	114 abc	274 ab	104 b
Buralin	101 bcd	265 ab	84 b
Butachlor	26 d	232 ab	37 b
Alachlor	48 cd	201 bc	48 b
Pendimenthalin	46 cd	200 bc	58 b
Thiobencarb	94 bcd	137 c	44 b
Làm cỏ bằng tay (2 lần)	182 a	289 a	230 a
Làm cỏ bằng tay (1 lần)	160 ab	263 ab	224 a
Không làm cỏ	75 cd	148 c	54 b
Trung bình	94	223	98

[a]Trung bình của bốn lần lặp lại. Sử dụng kiểm định Duncan ở mức ý nghĩa 5% để so sánh các trung bình ở cột . Giá trị $LSD_{0,05}$ để so sánh các trung bình chuẩn bị đất ở hàng là 73 kg/ha.

Bảng 12.11: Ảnh hưởng của ba nguồn urea và năm phương pháp làm đất trên chiều dài gié lúa. Giải thích sử dụng kiểm định Duncan thay vì LSD để nhấn mạnh nhân tố cần thiết.

Số lần cày bừa			Trung bình chiều dài gié, cm[a]		
Cày	Bừa	Cày xoay	Prilled	Bọc lưu huỳnh	Hạt cực nhỏ
1	1	0	20,8 a	21,6 a	22,1 a
1	3	0	19,6 b	21,2 a	21,9 a
1	1	1	20,5 b	20,4 b	22,4 a
1	2	2	20,9 a	20,1 a	21,3 a
2	2	0	21,6 a	21,7 a	21,2 a

[a]Trung bình của hai giống và bốn lần lặp lại. Sử dụng kiểm định Duncan ở mức ý nghĩa 5% để so sánh các trung bình ở hàng . Giá trị $LSD_{0,05}$ để so sánh các trung bình ở cột là 1,5 cm.

Bảng 12.12: Trọng lượng rơm của năm giống lúa khác nhau với các phương pháp xử lý rơm khác nhau: giải thích sử dụng kiểm định LSD cho cả hai nhân tố.

	Trọng lượng rơm, t/ha[a]					
Xử lý rơm	IR38	IR40	IR42	IR44	IR46	Trung bình
Loại bỏ rơm	3,44	5,00	3,56	3,72	3,80	3,90
Rải rơm	3,30	4,28	4,08	3,60	3,24	3,70
Đốt rơm	3,14	3,68	3,98	3,94	3,32	3,61
Ủ rơm	2,88	3,90	3,60	4,12	3,84	3,67
Trung bình	3,19	4,22	3,80	3,84	3,55	3,72

[a]Trung bình của năm lần lặp lại. Để so sánh trung bình ở cột, $LSD_{0,05}$ = 0,71 t/ha và ở hàng, $LSD_{0,05}$ = 0,65 t/ha.

Bảng 12.13: Ảnh hưởng của các nòi vi khuẩn gây bệnh cháy lá trên chiều dài vết bệnh của sáu giống lúa: giải thích sử dụng hai bộ chữ Duncan - từ a đến e để so sánh các trung bình ở cột và từ w đến z để so sánh các trung bình ở hàng.

	Chiều dài vết bệnh, cm[a]						
Nòi	IR8	IR20	IR1565	DV85	RD7	RĐ	T.B.
TB7803	23,5 b w	18,4 a x	4,9 ab y	5,3 ab y	18,9 b x	23,0 bc w	15,7
TB7805	16,0 c xy	15,0 b y	6,5 a z	6,2 a z	18,8 b wx	20,5 c w	13,8
TB7807	1,4 d w	1,4 f w	2,2 bc w	2,2 bc w	1,1 d w	1,0 d w	1,6
TB7808	30,2 a w	4,4 de y	2,7 bc y	3,2 abc y	5,5 c y	25,6 b x	11,9
TB7810	24,3 b x	5,3 de y	1,5 c z	3,2 abc yz	5,7 c y	30,0 a w	11,7
TB7814	23,4 b w	8,6 c y	3,6 abc z	4,8 abc z	18,3 b x	25,2 b w	14,0
TB7831	24,2 b x	2,7 ef y	1,9 bc y	3,0 abc y	4,8 c y	28,7 a w	10,9
TB7833	1,4 d w	1,2 f w	2,0 bc w	1,8 c w	2,6 cd w	2,2 d w	1,9
TB7841	29,0 a w	6,1 cd y	3,6 abc y	4,9 abc y	23,3 a x	30,4 a w	16,2
T.B.	19,3	7,0	3,2	3,8	11,0	20,7	10,8

[a]Trung bình của ba lần lặp lại. Sử dụng kiểm định Duncan ở mức ý nghĩa 5%.

Bảng 12.14: Chiều cao lúc thu hoạch của cỏ *Cyperus rotundus* bị ảnh hưởng bởi độ cao và chế độ nước: giải thích sự khác biệt cột cho nhân tố có hai mức độ.

	Chiều cao cây, cm[b]		
Chế độ nước[a]	Đất cao	Đất thấp	Khác biệt
Tháo nước tốt	17.50 a	34.58 c	- 17.08*
Bảo hoà	19.82 a	33.46 c	- 13.64[ns]
Bị ngập 7 NSKNM	17.82 a	84.60 a	- 66.68**
Bị ngập 14 NSKNM	24.85 a	85.07 a	- 60.22**
Bị ngập 21 NSKNM	23.82 a	71.60 ab	- 47.78**
Bị ngập 28 NSKNM	26.83 a	67.04 b	- 40.21**

[a]NSKNM = Ngày sau khi nhú mầm

[b]Trung bình của sáu lần lặp lại. Sử dụng kiểm định DUNCAN ở mức ý nghĩa 5% để so sánh các trung bình ở cột. ** = Ý nghĩa ở mức 1%, * = ý nghĩa ở mức 5%, [ns] = không ý nghĩa.

2. Đồ biểu hình thanh

Đối với số liệu của thí nghiệm thừa số trong đó không có nhân tố nào liên tục, đồ biểu hình thanh cho sự xen kẽ với sử dụng bảng. Ngoài các nguyên tắc đã được thảo luận cho thí nghiệm một nhân tố, sự lưu ý đầu tiên cho thí nghiệm thừa số là việc tạo chuỗi liên tục và sự gom nhóm của các nhân tố và các mức độ được trình bày. Ví dụ, thí nghiệm thừa số 2 x 2 gồm giống là nhân tố A và manganese dioxide (MnO_2) là nhân tố B, sẽ có bốn thanh - một tương ứng với một trong bốn tổ hợp nghiệm thức. Các thanh thường được sắp xếp thế nào để các mức độ của nhân tố quan tâm chính nằm kề lẫn nhau. Các mức độ của nhân tố kém quan trọng thì không cần. Do đó, nếu sự quan tâm đầu tiên của nhà nghiên cứu là để xác định ảnh hưởng của MnO_2, Hình 12.10a là cách trình bày thích hợp. Mặt khác, nếu nhà nghiên cứu muốn nhấn mạnh sự khác biệt giống, Hình 12.10b sẽ thích hợp hơn.

Bảng 12.15: Ảnh hưởng của các tổ hợp thuốc diệt cỏ khác nhau (W_1 đến W_6) và các phương pháp chuẩn bị đất (M_1 đến M_{10}) trên năng suất bắp: giải thích so sánh đúng giữa các trung bình khi tương tác giữa hai nhân tố không ý nghĩa.

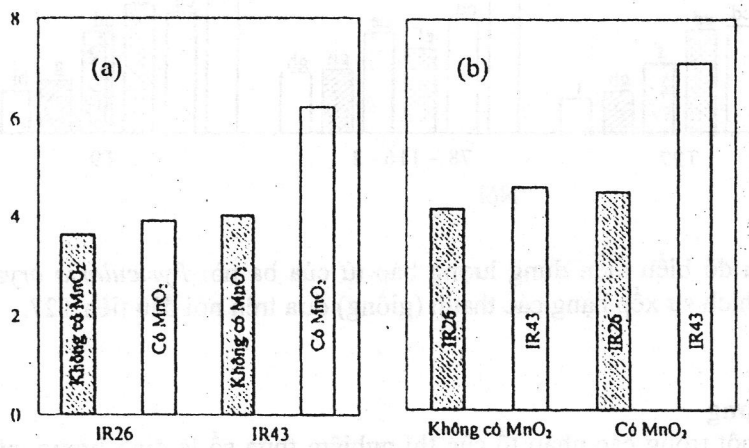
Chuẩn bị đất	Năng suất trung bình, t/haa						
	W_1	W_2	W_3	W_4	W_5	W_6	T.B.[b]
M_1	3,45	4,42	4,01	3,84	4,04	4,15	3,98 a
M_2	3,18	3,99	4,03	3,78	4,17	3,84	3,83 a
M_3	3,16	4,36	4,11	4,07	4,28	3,75	3,96 a
M_4	3,33	4,51	4,48	3,52	3,66	4,36	3,98 a
M_5	3,46	4,13	4,06	4,25	4,05	4,14	4,02 a
M_6	3,77	4,07	4,17	4,33	4,72	4,46	4,25 a
M_7	2,88	3,89	3,52	3,69	3,62	4,06	3,61 a
M_8	2,94	4,41	3,68	3,68	3,61	4,20	3,75 a
M_9	3,15	4,26	4,51	4,27	4,26	3,98	4,07 a
M_{10}	3,32	4,91	4,33	4,41	4,60	4,15	4,29 a
T.B.[b]	3,26 c	4,30 a	4,09 ab	3,98 a	4,10 ab	4,11 ab	

[a] Trung bình của ba lần lặp lại

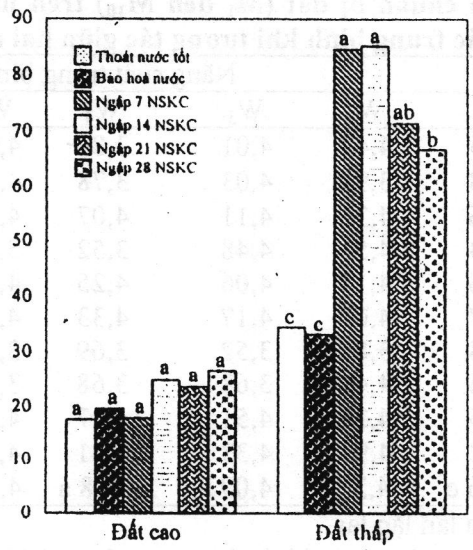
[b] Ở hàng (hoặc cột), các trung bình theo sau cùng chữ thì không khác biệt ở mức ý nghĩa 5%.

Khi một trong các nhân tố có hơn hai mức độ, sự tạo chuỗi liên tục đúng của các mức độ bên trong nhân tố cần được trình bày. Một vài ví dụ của sự tạo chuỗi liên tục:

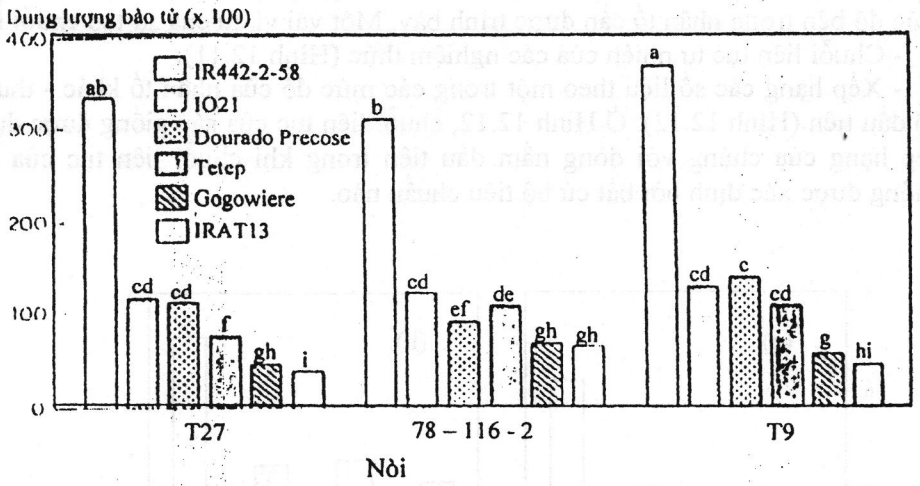
- Chuỗi liên tục tự nhiên của các nghiệm thức (Hình 12.11).
- Xếp hạng các số liệu theo một trong các mức độ của nhân tố khác - thường mức độ đầu tiên (Hình 12.12). Ở Hình 12.12, chuỗi liên tục của sáu giống được dựa trên sự xếp hạng của chúng với dòng nấm đầu tiên trong khi chuỗi liên tục của các dòng không được xác định bởi bất cứ bộ tiêu chuẩn nào.



Hình12.10: Biểu đồ hình thanh chỉ hai sự lựa chọn để gom nhóm các nhân tố: chọn (a) để nhấn mạnh việc áp dụng MnO_2 và chọn (b) để nhấn mạnh sự khác biệt của giống



Hình12.11: Biểu đồ hình thanh chỉ sự sắp xếp liên tiếp các thanh theo bản chất của nghiệm thức (các mức độ trong một nhân tố). Các trung bình với kiểu canh tác lúa giống nhau (cao hoặc thấp) có cùng chữ thì khác biệt không ý nghĩa ở mức 5%.



Hình12.12: Biểu đồ biểu diễn dung lượng bào tử của ba nòi *Pyricularia oryzae* trên sáu giống: giải thích sự xếp hạng các thanh (giống) dựa trên nòi đầu tiên T27.

3. Đồ thị đường

Khi ít nhất một trong các nhân tố của thí nghiệm thừa số là định lượng, nên dùng đồ thị đường. Sự quan tâm và các nguyên tắc cơ bản để sử dụng đồ thị đường, mô tả cho thí nghiệm một nhân tố cũng có thể áp dụng cho thí nghiệm thừa số. Một vài hướng dẫn bổ sung:

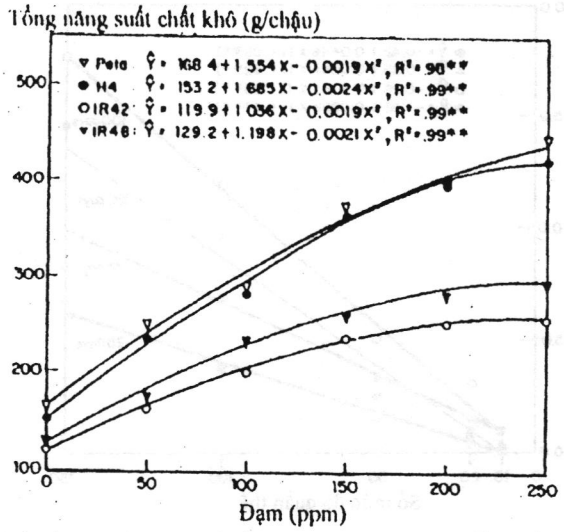
(1) Đối với thí nghiệm thừa số A x b mà nhân tố A là định lượng và nhân tố B riêng biệt, đồ thị đường.sẽ có các mức độ của nhân tố định lượng A trên trục X, phản ứng trên trục Y, và vẽ một đường cho mỗi mức độ của nhân tố B riêng biệt (Hình 12.13).

Ở Hình 12.13, có hai nhân tố: giống có bốn mức độ và phân đạm có sáu mức độ. Vì mức độ đạm là biến định lượng, nên đường hồi qui để trình bày phản ứng đối với đạm thì thích hợp cho một trong bốn giống, Mặt khác, nếu không ước lượng phương trình hồi qui, có thể vẽ đường thẳng nối các điểm kế cận (Hình 12.14).

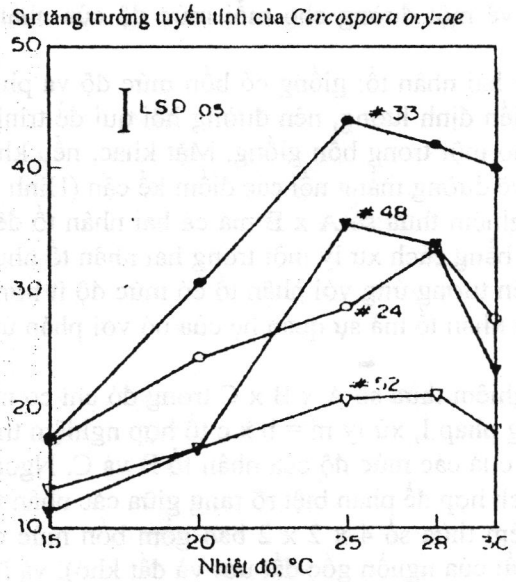
(2) Đối với thí nghiệm thừa số A x B mà cả hai nhân tố đều định lượng, nên tuân theo phương pháp 1, bằng cách xử lý một trong hai nhân tố như biến rời, nhân tố được xử lý như biến rời nên tương ứng với nhân tố có mức độ ít hơn và tầm quan trọng của nó thấp hơn hoặc với nhân tố mà sự quan hệ của nó với phản ứng của cây trồng không rõ (Hình 12.15).

(3) Đối với thí nghiệm thừa số A x B x C trong đó chỉ có nhân tố A là định lượng, nên tuân theo phương pháp I, xử lý m = b x c tổ hợp nghiệm thức như các mức độ của nhân tố rời, với b và c là các mức độ của nhân tố B và C. Ngoài ra, việc sử dụng cách nhận dạng đường thích hợp để phân biệt rõ ràng giữa các nhân tố được khảo sát. Ví dụ, kết quả của thí nghiệm thừa số 4 x 2 x 2 bao gồm bốn mức độ tiêm chủng trong hệ thống, hai loại đất (đất của nguồn gốc đất ướt và đất khô), và hai chế độ nước (khô và bị ngập) như được chỉ ở Hình 12.16: với đường liên tục cho nguồn gốc đất khô và đường đứt quảng cho nguồn gốc đất ướt, trong khi các điểm tô đen (• và ▼) liên quan đến điều kiện khô và những điểm rỗng (o và ∇) liên quan đến điều kiện bị ngập.

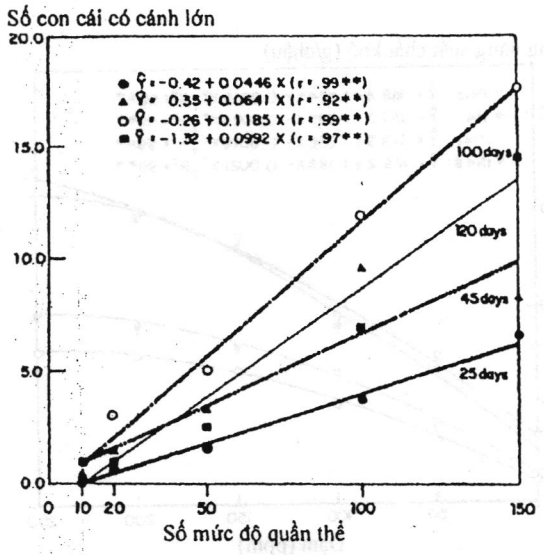
(4) Đối với thí nghiệm có từ ba nhân tố trở lên, sử dụng nhiều hơn một đồ thị, nếu cần. Nhân tố, mà trong đó mỗi mức độ được trình bày bằng đồ thị, thường là một nhân tố mà ở đó ảnh hưởng chính không được quan tâm chủ yếu, hoặc ảnh hưởng của nó thì lớn (Hình 12.17).



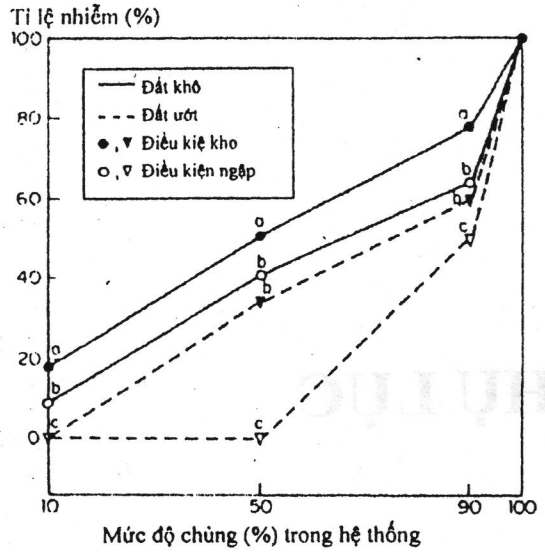
Hình 12.13: Phản ứng đối với đạm của bốn giống lúa: giải thích cách sử dụng đồ thị đường cho thí nghiệm thừa số, trong đó một nhân tố là định lượng, và phương trình hồi qui phù hợp.



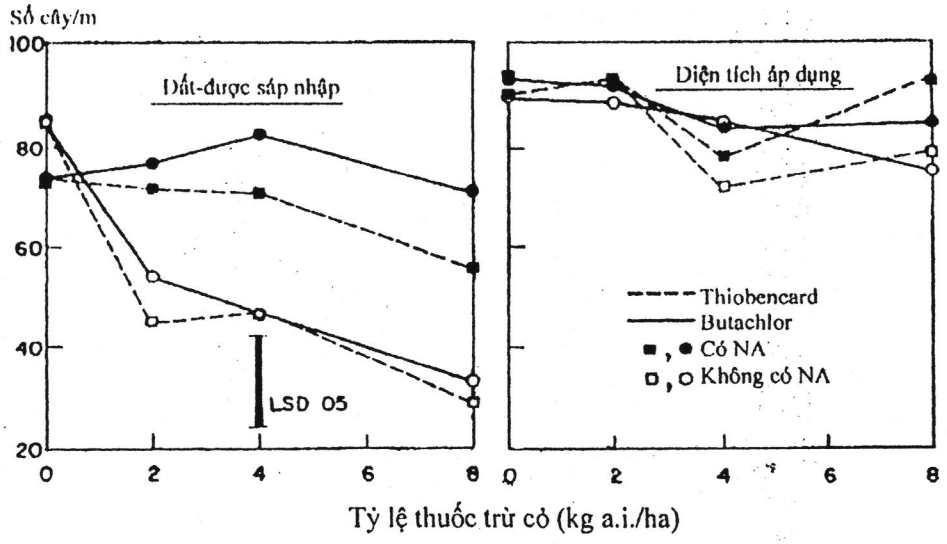
Hình 12.14: Sự tăng trưởng tuyến tính của của bốn nòi *Cercospora oryzae* bị ảnh hưởng bởi nhiệt độ: giải thích cách sử dụng đồ thị đường cho thí nghiệm thừa số, trong đó một nhân tố là định lượng và phương trình hồi qui không phù hợp.



Hình12.15: Khả năng nhiễm của rơm được vùi trong đất của nguồn gốc đất ướt và đất khô: giải thích cách sử dụng sự nhận dạng đường thích hợp đối với các nhân tố phân biệt, tách biệt các trung bình ở mỗi mức tiêm chủng bằng kiểm định Duncan ở mức 5%.



Hình12.16: Ảnh hưởng của tỷ lệ, loại và phương pháp áp dụng thuốc trừ cỏ và naphthalic anhydride (NA) trên sự nảy mầm của lúa hạt được sấy khô: giải thích cách sử dụng nhiều hơn một đồ thị cho số liệu của thí nghiệm thừa số.



Hình12.17: Ảnh hưởng của tỷ lệ, loại và phương pháp áp dụng thuốc trừ cỏ và naphthalic anhydride (NA) trên sự nảy mầm của lúa hạt được sấy khô: giải thích cách sử dụng nhiều hơn một đồ thị cho số liệu của thí nghiệm thừa số.

PHỤ LỤC

Phụ lục I: TUNG ĐỘ CỦA ĐƯỜNG CONG CHUẨN TẮC

z	Số thập phân thứ hai của z									
	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	.3989	.3989	.3989	.3988	.3986	.3984	.3982	.3980	.3977	.3973
0.1	.3970	.3965	.3961	.3956	.3951	.3945	.3939	.3932	.3925	.3918
0.2	.3910	.3902	.3894	.3885	.3876	.3867	.3857	.3847	.3836	.3825
0.3	.3814	.3802	.3790	.3778	.3765	.3752	.3739	.3725	.3712	.3697
0.4	.3683	.3668	.3653	.3637	.3621	.3605	.3589	.3572	.3555	.3536
0.5	.3521	.3503	.3485	.3467	.3448	.3429	.3410	.3391	.3372	.3352
0.6	.3332	.3312	.3292	.3271	.3251	.3230	.3209	.3187	.3166	.3144
0.7	.3123	.3101	.3079	.3056	.3034	.3011	.2989	.2966	.2943	.2920
0.8	.2897	.2874	.2850	.2827	.2803	.2780	.2756	.2732	.2709	.2685
0.9	.2661	.2637	.2613	.2589	.2565	.2541	.2516	.2492	.2468	.2444
1.0	.2420	.2396	.2371	.2347	.2323	.2299	.2275	.2251	.2227	.2203
1.1	.2179	.2155	.2131	.2107	.2083	.2059	.2036	.2012	.1989	.1965
1.2	.1942	.1919	.1895	.1872	.1849	.1826	.1804	.1781	.1758	.1736
1.3	.1714	.1691	.1669	.1647	.1626	.1604	.1582	.1561	.1539	.1518
1.4	.1497	.1476	.1456	.1435	.1415	.1394	.1374	.1354	.1334	.1315
1.5	.1295	.1276	.1257	.1238	.1219	.1200	.1182	.1163	.1145	.1127
1.6	.1109	.1092	.1074	.1057	.1040	.1023	.1006	.0989	.0973	.0957
1.7	.0940	.0925	.0909	.0893	.0878	.0863	.0848	.0833	.0818	.0804
1.8	.0790	.0775	.0761	.0748	.0734	.0721	.0707	.0694	.0681	.0669
1.9	.0656	.0664	.0632	.0620	.0608	.0596	.0584	.0573	.0562	.0551
2.0	.0540	.0529	.0519	.0508	.0498	.0488	.0478	.0468	.0459	.0449
2.1	.0440	.0431	.0422	.0413	.0404	.0396	.0387	.0379	.0371	.0363
2.2	.0355	.0347	.0339	.0332	.0325	.0317	.0310	.0303	.0297	.0290
2.3	.0283	.0277	.0270	.0264	.0258	.0252	.0246	.0241	.0235	.0229
2.4	.0224	.0219	.0213	.0208	.0203	.0198	.0194	.0189	.0184	.0180
2.5	.0175	.0171	.0167	.0163	.0158	.0154	.0151	.0147	.0143	.0139
2.6	.0136	.0132	.0129	.0126	.0122	.0119	.0116	.0113	.0110	.0107
2.7	.0104	.0101	.0099	.0093	.0091	.0090	.0088	.0086	.0084	.0081
2.8	.0079	.0077	.0075	.0073	.0071	.0069	.0067	.0065	.0063	.0061
2.9	.0060	.0058	.0056	.0055	.0053	.0051	.0050	.0048	.0047	.0046
	Số thập phân thứ nhất của z									
	0.0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
3	.0044	.0033	.0024	.0017	.0012	.0009	.0006	.0004	.0003	.0002
4	.0001	.0001	.0001	.0000	.0000	.0000	.0000	.0000	.0000	.0000

Phụ lục II: BẢNG PHÂN BỐ Z
(Diện tích dưới đường cong chuẩn tắc tính từ 0 đến z)

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.0000	0.0040	0.0080	0.0120	0.0160	0.0199	0.0239	0.0279	0.0319	0.0359
0.1	.0398	.0438	.0478	.0517	.0557	.0596	.0636	.0675	.0714	0753
0.2	.0793	.0832	.0871	.0910	.0948	.0987	.1026	.1064	.1103	.1141
0.3	.1179	.1217	.1255	.1293	.1331	.1368	.1406	.1443	.1480	.1517
0.4	.1554	.1591	.1628	.1664	.1700	.1736	.1772	.1808	.1844	.1879
0.5	.1915	.1950	.1985	.2019	.2054	.2088	.2123	.2157	.2190	.2224
0.6	.2257	.2291	.2324	.2357	.2389	.2422	.2454	.2486	.2517	2549
0.7	.2580	.2611	.2642	.2673	.2704	.2734	.2764	.2794	.2823	.2852
0.8	.2881	.2910	.2939	.2967	.2995	.3023	.3051	.3087	.3106	3133
0.9	.3159	.3186	.3212	.3238	.3264	.3289	.3315	.3340	.3365	.3389
1.0	.3413	.3438	.3461	.3485	.3508	.3531	.3554	.3577	.3599	3621
1.1	.3643	.3665	.3686	.3708	.3729	.3749	.3770	.3790	.3810	3830
1.2	.3849	.3869	.3888	.3907	.3925	.3944	.3962	.3980	.3997	4015
1.3	.4032	.4049	.4066	.4082	.4099	.4115	.4131	.4147	.4162	4177
1.4	.4192	.4207	.4222	.4236	.4251	.4265	.4279	.4292	.4306	4319
1.5	.4332	.4345	.4357	.4370	.4382	.4394	.4406	.4418	.4429	.4441
1.6	.4452	.4463	.4474	.4484	.4495	.4505	.4515	.4525	4535	4545
1.7	.4554	.4564	.4573	.4582	.4591	.4599	.4608	.4616	4625	4633
1.8	.4641	.4649	.4656	.4664	.4671	.4678	.4686	.4693	.4699	4706
1.9	.4713	.4719	.4726	.4732	.4738	.4744	.4750	.4756	.4761	.4767
2.0	.4772	.4778	.4783	.4788	.4793	.4798	.4803	.4808	.4812	4817
2.1	.4821	.4826	.4830	.4834	.4838	.4842	.4846	.4850	.4854	4857
2.2	.4861	.4864	.4868	.4871	.4875	.4878	.4881	.4884	.4887	4890
2.3	.4893	.4896	.4898	.4901	.4904	.4906	.4909	.4911	.4913	4916
2.4	.4918	.4920	.4922	.4925	.4927	.4929	.4931	.4932	.4934	.4936
2.5	.4938	.4940	.4941	.4943	.4945	.4946	.4948	.4949	.4951	.4952
2.6	.4953	.4955	.4956	.4957	.4959	.4960	.4961	.4962	.4963	.4964
2.7	.4965	.4966	.4967	.4968	.4969	.4970	.4971	.4972	4973	4974
2.8	.4974	.4975	.4976	.4977	.4977	.4978	.4979	.4979	.4980	.4981
2.9	.4981	.4982	.4982	.4983	.4984	.4984	.4985	.4985	.4986	.4986
3.0	.4987	.4987	.4987	.4988	.4988	.4989	.4989	.4989	.4990	4990
3.1	.4990	.4991	.4991	.4991	4992	.4992	.4992	.4992	.4993	4993
3.2	.4993	.4993	.4994	.4994	.4994	.4994	.4994	.4995	.4995	4995
3.3	.4995	.4995	.4995	.4996	.4996	.4996	.4996	.4996	.4996	.4997
3.4	.4997	.4997	.4997	.4997	.4997	.4997	.4997	.4997	.4997	.4998
3.6	.4998	.4998	.4999	.4999	.4999	.4999	.4999	.4999	.4999	.4999
3.9	.6000									

Phụ lục III: Bảng giá trị $e^{-\lambda}$

$(0 < \lambda < 1)$

λ	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	1.0000	.9900	.9802	.9704	.9608	.9512	.9418	.9324	.9231	.9139
0.1	.9048	.8958	.8869	.8781	.8694	.8607	.8521	.8437	.8353	.8270
0.2	.8187	.8106	.8025	.7945	.7866	.7788	.7711	.7634	.7558	.7483
0.3	.7408	.7334	.7261	.7189	.7118	.7047	.6977	.6907	.6839	.6771
0.4	.6703	.6636	.6570	.6505	.6440	.6376	.6313	.6250	.6188	.6126
0.5	.6065	.6005	.5945	.5886	.5827	.5770	.5712	.5655	.5599	.5543
0.6	.5488	.5434	.5379	.5326	.5273	.5220	.5169	.5117	.5066	.5016
0.7	.4966	.4916	.4868	.4819	.4771	.4724	.4677	.4630	.4584	.4538
0.8	.4493	.4449	.4404	.4360	.4317	.4274	.4232	.4190	.4148	.4107
0.9	.4066	.4025	.3985	.3946	.3906	.3867	.3829	.3791	.3753	.3716

$(\lambda = 1, 2, 3, \ldots, 10)$

λ	1	2	3	4	5	6	7	8	9	10
$e^{-\lambda}$	.36788	.13534	.04979	.01832	.006738	.002479	.000912	.000335	.000123	.000045

Chú thích: Để đạt được các giá trị $e^{-\lambda}$ cho những giá trị λ khác, sử dụng luật lũy thừa.
Ví dụ: $e^{-3,48} = (e^{-3,00})(e^{-0,48}) = (0,04979)(0,6188) = 0,03081$

Phụ lục IV: BẢNG PHÂN BỐ t (2 đuôi)

df-1	.9	8	7	6	.5	.4	.3	2	.1	.05	.02	.01	.001
1	.158	.325	.510	.727	1.000	1 376	1.963	3.078	6.314	12.706	31.821	63.657	636.619
2	142	.289	.445	617	.816	1.061	1.386	1 886	2 920	4.303	6.965	9.925	31.598
3	.137	277	.424	.584	.765	.978	1.250	1.638	2.353	3.182	4 541	5.841	12.924
4	.134	.271	.414	.569	.741	.941	1.190	1 533	2.132	2.776	3 747	4.604	8.610
5	.132	.267	.408	.559	727	.920	1.156	1.476	2 015	2.571	3.365	4.032	6.869
6	131	.265	.404	.553	.718	.906	1 134	1.440	1 943	2.447	3.143	3 707	5.959
7	.130	.263	.402	.549	711	.896	1.119	1.415	1 895	2.365	2.998	3 499	5.408
8	.130	.262	.399	.546	.706	.889	1 108	1.397	1.860	2.306	2.896	3.355	5.041
9	.129	.261	.398	.543	.703	.883	1.100	1.383	1.833	2.262	2.821	3 250	4.781
10	.129	.260	.397	542	.700	.879	1.093	1.372	1 812	2.228	2.764	3 169	4 587
11	129	.260	396	540	.697	.876	1.088	1.363	1.796	2.201	2.718	3.106	4.437
12	.128	.259	.395	.539	.695	.873	1 083	1.356	1.782	2.179	2.681	3 055	4.318
13	.128	.259	.394	.538	.694	.870	1 079	1.350	1.771	2.160	2.650	3 012	4.221
14	.128	.258	.393	537	.692	.868	1.076	1 345	1.761	2.145	2.624	2.977	4 140
15	.128	.258	.393	536	.691	866	1.074	1 341	1.753	2.131	2 602	2 947	4.073
16	.128	.258	.392	.535	690	.865	1.071	1 337	1.746	2.120	2 583	2 921	4.015
17	.128	.257	392	.534	689	863	1.069	1.333	1.740	2 110	2.567	2 898	3 965
18	127	.257	.392	.534	688	.862	1.067	1 330	1.734	2.101	2 552	2.878	3.922
19	.127	.257	.391	.533	.688	.861	1 066	1.328	1.729	2.093	2 539	2.861	3.883
20	.127	.257	.391	.533	687	.860	1.064	1 325	1.725	2.086	2.528	2.845	3.850
21	.127	.257	.391	.532	.686	.859	1.063	1 323	1.721	2 080	2 518	2 831	3 819
22	.127	.256	.390	.532	686	858	1.061	1.321	1.717	2.074	2 508	2 819	3.792
23	127	.256	.390	.532	685	858	1 060	1.319	1.714	2 069	2.500	2 807	3.767
24	.127	.256	.390	.531	685	.857	1 059	1.318	1.711	2 064	2 492	2 797	3 745
25	127	.256	.390	.531	684	856	1.058	1.316	1.708	2.060	2 485	2 787	3 725
26	.127	256	390	.531	684	.856	1 058	1.315	1.706	2 056	2.479	2.779	3.707
27	.127	.256	.389	.531	684	.855	1.057	1.314	1 703	2.052	2 473	2 771	3 690
28	.127	.256	.389	.530	683	855	1 056	1.313	1 701	2.048	2.467	2 763	3 674
29	.127	256	.389	530	683	854	1 055	1 311	1.699	2 045	2 462	2 756	3 659
30	.127	256	.389	.530	683	.854	1 055	1.310	1.697	2 042	2.457	2.750	3 646
40	126	255	388	.529	681	.851	1 050	1.303	1 684	2 021	2.423	2 704	3 551
60	.126	.254	.387	.527	679	.848	1.046	1 296	1.671	2 000	2.390	2.660	3.460
120	.126	.254	.386	.526	.677	.845	1.041	1.289	1.658	1.980	2.358	2.617	3.373
∞	.126	.253	.385	.524	.674	.842	1.036	1.282	1.645	1.960	2.326	2.576	3.291

Phụ lục V : BẢNG PHÂN BỐ CHI BÌNH PHƯƠNG (χ^2)

df \ p	.995	.99	.975	.95	.90	.75	.50	.25	.10	.05	.025	.01	.005
1	.0000	.0002	.0010	.0039	.0158	.102	.455	1.32	2.71	3.84	5.02	6.63	7.88
2	.0100	.0201	.0506	.103	.211	.575	1.39	2.77	4.61	5.99	7.38	9.21	10.6
3	.072	.115	.216	.352	.584	1.21	2.37	4.11	6.25	7.81	9.35	11.3	12.8
4	.207	.297	.484	.711	1.06	1.92	3.36	5.39	7.78	9.49	11.1	13.3	14.9
5	.412	.554	.831	1.15	1.61	2.67	4.35	6.63	9.24	11.1	12.8	15.1	16.7
6	.676	.872	1.24	1.64	2.20	3.45	5.35	7.84	10.6	12.6	14.4	16.8	18.5
7	.989	1.24	1.69	2.17	2.83	4.25	6.35	9.04	12.0	14.1	16.0	18.5	20.3
8	1.34	1.65	2.18	2.73	3.49	5.07	7.34	10.2	13.4	15.5	17.5	20.1	22.0
9	1.73	2.09	2.70	3.33	4.17	5.90	8.34	11.4	14.7	16.9	19.0	21.7	23.6
10	2.16	2.56	3.25	3.94	4.87	6.74	9.34	12.5	16.0	18.3	20.5	23.2	25.2
11	2.60	3.05	3.82	4.57	5.58	7.58	10.3	13.7	17.3	19.7	21.9	24.7	26.3
12	3.07	3.57	4.40	5.23	6.30	8.44	11.3	14.8	18.5	21.0	23.3	26.2	28.3
13	3.57	4.11	5.01	5.89	7.04	9.30	12.3	16.0	19.8	22.4	24.7	27.7	29.8
14	4.07	4.66	5.63	6.57	7.79	10.2	13.3	17.1	21.1	23.7	26.1	29.1	31.3
15	4.60	5.23	6.26	7.26	8.55	11.0	14.3	18.2	22.3	25.0	27.5	30.6	32.8
16	5.14	5.81	6.91	7.96	9.31	11.9	15.3	19.4	23.5	26.3	28.8	32.0	34.3
17	5.70	6.41	7.56	8.67	10.1	12.8	16.3	20.5	24.8	27.6	30.2	33.4	35.7
18	6.26	7.01	8.23	9.39	10.9	13.7	17.3	21.6	26.0	28.9	31.5	34.8	37.2
19	6.84	7.63	8.91	10.1	11.7	14.6	18.3	22.7	27.2	30.1	32.9	36.2	38.6
20	7.43	8.26	9.59	10.9	12.4	15.5	19.3	23.8	28.4	31.4	34.2	37.6	40.0
21	8.03	8.90	10.3	11.6	13.2	16.3	20.3	24.9	29.6	32.7	35.5	38.9	41.4
22	8.64	9.54	11.0	12.3	14.0	17.2	21.3	26.0	30.8	33.9	36.8	40.3	42.8
23	9.26	10.2	11.7	13.1	14.8	18.1	22.3	27.1	32.0	35.2	38.1	41.6	44.2
24	9.89	10.9	12.4	13.8	15.7	19.0	23.3	28.2	33.2	36.4	39.4	43.0	45.6
25	10.5	11.5	13.1	14.6	16.5	19.9	24.3	29.3	34.4	37.7	40.6	44.3	46.9
26	11.2	12.2	13.8	15.4	17.3	20.8	25.3	30.4	35.6	38.9	41.9	45.6	48.3
27	11.8	12.9	14.6	16.2	18.1	21.7	26.3	31.5	36.7	40.1	43.2	47.0	49.6
28	12.5	13.6	15.3	16.9	18.9	22.7	27.3	32.6	37.9	41.3	44.5	48.3	51.0
29	13.1	14.3	16.0	17.7	19.8	23.6	28.3	33.7	39.1	42.6	45.7	49.6	52.3
30	13.8	15.0	16.8	18.5	20.6	24.5	29.3	34.8	40.3	43.8	47.0	50.9	53.7
40	20.7	22.2	24.4	26.5	29.1	33.7	39.3	45.6	51.8	55.8	59.3	63.7	66.8
50	28.0	29.7	32.4	34.8	37.7	42.9	49.3	56.3	63.2	67.5	71.4	76.2	79.5
60	35.5	37.5	40.5	43.2	46.5	52.3	59.3	67.0	74.4	79.1	83.3	88.4	92.0
70	43.3	45.4	48.8	51.7	55.3	61.7	69.3	77.6	85.5	90.5	95.0	100.4	104.2
80	51.2	53.5	57.2	60.4	64.3	71.1	79.3	88.1	96.6	101.9	106.6	112.3	116.3
90	59.2	61.8	65.6	69.1	73.3	80.6	89.3	98.6	107.6	113.1	118.1	124.1	128.3
100	67.3	70.1	74.2	77.9	82.4	90.1	99.3	109.1	118.5	124.3	129.6	135.8	140.2

Phụ lục VI: BẢNG PHÂN BỐ F (5% in lợt và 1% in đậm)

df_2 (Độ tự do của mẫu số)	df_1 (Độ tự do của tử số)																								df_2
	1	2	3	4	5	6	7	8	9	10	11	12	14	16	20	24	30	40	50	75	100	200	500	∞	
1	161	200	216	225	230	234	237	239	241	242	243	244	245	246	248	249	250	251	252	253	253	254	254	254	1
	4052	**4999**	**5403**	**5625**	**5764**	**5859**	**5928**	**5981**	**6022**	**6056**	**6082**	**6106**	**6142**	**6169**	**6208**	**6234**	**6261**	**6286**	**6302**	**6323**	**6334**	**6352**	**6361**	**6366**	
2	18.51	19.00	19.16	19.25	19.30	19.33	19.36	19.37	19.38	19.39	19.40	19.41	19.42	19.43	19.44	19.45	19.46	19.47	19.47	19.48	19.49	19.49	19.50	19.50	2
	98.49	**99.00**	**99.17**	**99.25**	**99.30**	**99.33**	**99.36**	**99.37**	**99.39**	**99.40**	**99.41**	**99.42**	**99.43**	**99.44**	**99.45**	**99.46**	**99.47**	**99.48**	**99.48**	**99.49**	**99.49**	**99.49**	**99.50**	**99.50**	
3	10.13	9.55	9.28	9.12	9.01	8.94	8.88	8.84	8.81	8.78	8.76	8.74	8.71	8.69	8.66	8.64	8.62	8.60	8.58	8.57	8.56	8.54	8.54	8.53	3
	34.12	**30.82**	**29.46**	**28.71**	**28.24**	**27.91**	**27.67**	**27.49**	**27.34**	**27.23**	**27.13**	**27.05**	**26.92**	**26.83**	**26.69**	**26.60**	**26.50**	**26.41**	**26.35**	**26.27**	**26.23**	**26.18**	**26.14**	**26.12**	
4	7.71	6.94	6.59	6.39	6.26	6.16	6.09	6.04	6.00	5.96	5.93	5.91	5.87	5.84	5.80	5.77	5.74	5.71	5.70	5.68	5.66	5.65	5.64	5.63	4
	21.20	**18.00**	**16.69**	**15.98**	**15.52**	**15.21**	**14.98**	**14.80**	**14.66**	**14.54**	**14.46**	**14.37**	**14.24**	**14.15**	**14.02**	**13.93**	**13.83**	**13.74**	**13.69**	**13.61**	**13.57**	**13.52**	**13.48**	**13.46**	
5	6.61	5.79	5.41	5.19	5.05	4.95	4.88	4.82	4.78	4.74	4.70	4.68	4.64	4.60	4.56	4.53	4.50	4.46	4.44	4.42	4.40	4.38	4.37	4.36	5
	16.26	**13.27**	**12.06**	**11.39**	**10.97**	**10.67**	**10.45**	**10.29**	**10.15**	**10.05**	**9.96**	**9.89**	**9.77**	**9.68**	**9.55**	**9.47**	**9.38**	**9.29**	**9.24**	**9.17**	**9.13**	**9.07**	**9.04**	**9.02**	
6	5.99	5.14	4.76	4.53	4.39	4.28	4.21	4.15	4.10	4.06	4.03	4.00	3.96	3.92	3.87	3.84	3.81	3.77	3.75	3.72	3.71	3.69	3.68	3.67	6
	13.74	**10.92**	**9.78**	**9.15**	**8.75**	**8.47**	**8.26**	**8.10**	**7.98**	**7.87**	**7.79**	**7.72**	**7.60**	**7.52**	**7.39**	**7.31**	**7.23**	**7.14**	**7.09**	**7.02**	**6.99**	**6.94**	**6.90**	**6.88**	
7	5.59	4.74	4.35	4.12	3.97	3.87	3.79	3.73	3.68	3.63	3.60	3.57	3.52	3.49	3.44	3.41	3.38	3.34	3.32	3.29	3.28	3.25	3.24	3.23	7
	12.25	**9.55**	**8.45**	**7.85**	**7.46**	**7.19**	**7.00**	**6.84**	**6.71**	**6.62**	**6.54**	**6.47**	**6.36**	**6.27**	**6.15**	**6.07**	**5.98**	**5.90**	**5.85**	**5.78**	**5.75**	**5.70**	**5.67**	**5.65**	
8	5.32	4.46	4.07	3.84	3.69	3.58	3.50	3.44	3.39	3.34	3.31	3.28	3.23	3.20	3.15	3.12	3.08	3.05	3.03	3.00	2.98	2.96	2.94	2.93	8
	11.26	**8.65**	**7.59**	**7.01**	**6.63**	**6.37**	**6.19**	**6.03**	**5.91**	**5.82**	**5.74**	**5.67**	**5.56**	**5.48**	**5.38**	**5.28**	**5.20**	**5.11**	**5.06**	**5.00**	**4.96**	**4.91**	**4.88**	**4.86**	
9	5.12	4.26	3.86	3.63	3.48	3.37	3.29	3.23	3.18	3.13	3.10	3.07	3.02	2.98	2.93	2.90	2.86	2.82	2.80	2.77	2.76	2.73	2.72	2.71	9
	10.56	**8.02**	**6.99**	**6.42**	**6.06**	**5.80**	**5.62**	**5.47**	**5.36**	**5.26**	**5.18**	**5.11**	**5.00**	**4.92**	**4.80**	**4.73**	**4.64**	**4.56**	**4.51**	**4.45**	**4.41**	**4.36**	**4.33**	**4.31**	
10	4.96	4.10	3.71	3.48	3.33	3.22	3.14	3.07	3.02	2.97	2.94	2.91	2.86	2.82	2.77	2.74	2.70	2.67	2.64	2.61	2.59	2.56	2.55	2.54	10
	10.04	**7.56**	**6.55**	**5.99**	**5.64**	**5.39**	**6.21**	**5.06**	**4.96**	**4.86**	**4.78**	**4.71**	**4.60**	**4.52**	**4.41**	**4.33**	**4.25**	**4.17**	**4.12**	**4.05**	**4.01**	**3.96**	**3.93**	**3.91**	
11	4.84	3.98	3.59	3.36	3.20	3.09	3.01	2.95	2.90	2.86	2.82	2.79	2.74	2.70	2.65	2.61	2.57	2.53	2.50	2.47	2.45	2.42	2.41	2.40	11
	9.65	**7.20**	**6.22**	**5.67**	**5.32**	**5.07**	**4.88**	**4.74**	**4.63**	**4.54**	**4.46**	**4.40**	**4.29**	**4.21**	**4.10**	**4.02**	**3.94**	**3.86**	**3.80**	**3.74**	**3.70**	**3.66**	**3.62**	**3.60**	
12	4.75	3.88	3.49	3.26	3.11	3.00	2.92	2.85	2.80	2.76	2.72	2.69	2.64	2.60	2.54	2.50	2.46	2.42	2.40	2.36	2.35	2.32	2.31	2.30	12
	9.33	**6.93**	**5.95**	**5.41**	**5.06**	**4.82**	**4.65**	**4.50**	**4.39**	**4.30**	**4.22**	**4.16**	**4.05**	**3.98**	**3.86**	**3.78**	**3.70**	**3.61**	**3.56**	**3.49**	**3.46**	**3.41**	**3.38**	**3.36**	
13	4.67	3.80	3.41	3.18	3.02	2.92	2.84	2.77	2.72	2.67	2.63	2.60	2.55	2.51	2.46	2.42	2.38	2.34	2.32	2.28	2.26	2.24	2.22	2.21	13
	9.07	**6.70**	**5.74**	**5.20**	**4.86**	**4.62**	**4.44**	**4.30**	**4.19**	**4.10**	**4.02**	**3.96**	**3.85**	**3.78**	**3.67**	**3.59**	**3.51**	**3.42**	**3.37**	**3.30**	**3.27**	**3.21**	**3.18**	**3.16**	

Phụ lục VI: BẢNG PHÂN BỐ F (tiếp theo)

df_2 (Độ tự do của mẫu số)	df_2 (Độ tự do của tử số)																								
	1	2	3	4	5	6	7	8	9	10	11	12	14	16	20	24	30	40	50	75	100	200	500	∞	
14	4.60	3.74	3.34	3.11	2.96	2.85	2.77	2.70	2.65	2.60	2.56	2.53	2.48	2.44	2.39	2.35	2.31	2.27	2.24	2.21	2.19	2.16	2.14	2.13	14
	8.86	6.51	5.56	5.03	4.69	4.46	4.28	4.14	4.03	3.94	3.86	3.80	3.70	3.62	3.51	3.43	3.34	3.26	3.21	3.14	3.11	3.06	3.02	3.00	
15	4.54	3.68	3.29	3.06	2.90	2.79	2.70	2.64	2.59	2.55	2.51	2.48	2.43	2.39	2.33	2.29	2.25	2.21	2.18	2.15	2.12	2.10	2.08	2.07	15
	8.68	6.36	5.42	4.89	4.56	4.32	4.14	4.00	3.89	3.80	3.73	3.67	3.56	3.48	3.36	3.29	3.20	3.12	3.07	3.00	2.97	2.92	2.89	2.87	
16	4.49	3.63	3.24	3.01	2.85	2.74	2.66	2.59	2.54	2.49	2.45	2.42	2.37	2.33	2.28	2.24	2.20	2.16	2.13	2.09	2.07	2.04	2.02	2.01	16
	8.53	6.23	5.29	4.77	4.44	4.20	4.03	3.89	3.78	3.69	3.61	3.55	3.45	3.37	3.25	3.18	3.10	3.01	2.96	2.98	2.86	2.80	2.77	2.75	
17	4.45	3.59	3.20	2.96	2.81	2.70	2.62	2.55	2.50	2.45	2.41	2.38	2.33	2.29	2.23	2.19	2.15	2.11	2.08	2.04	2.02	1.99	1.97	1.96	17
	8.40	6.11	5.18	4.67	4.34	4.10	3.93	3.79	3.68	3.59	3.52	3.45	3.35	3.27	3.16	3.08	3.00	2.92	2.86	2.79	2.76	2.70	2.67	2.65	
18	4.41	3.55	3.16	2.93	2.77	3.66	2.58	2.51	2.46	2.41	2.37	2.34	2.29	2.25	2.19	2.15	2.11	2.07	2.04	2.00	1.98	1.95	1.93	1.92	18
	8.28	6.01	5.09	4.58	4.25	4.01	3.85	3.71	3.60	3.51	3.44	3.37	3.27	3.19	3.07	3.00	2.91	2.83	2.78	2.71	2.68	2.62	2.59	2.57	
19	4.38	3.52	3.13	2.90	2.74	2.63	2.55	2.48	2.43	2.38	2.34	2.31	2.26	2.21	2.15	2.11	2.07	2.02	2.00	1.96	1.94	1.91	1.90	1.88	19
	8.18	5.93	5.01	4.50	4.17	3.94	3.77	3.63	3.52	3.43	3.36	3.30	3.19	3.12	3.00	2.92	2.84	2.76	2.70	2.63	2.60	2.54	2.51	2.49	
20	4.35	3.49	3.10	2.87	2.71	2.60	2.52	2.45	2.40	2.35	2.31	2.28	2.23	2.18	2.12	2.08	2.04	1.99	1.96	1.92	1.90	1.87	1.85	1.84	20
	8.10	5.85	4.94	4.43	4.10	3.87	3.71	3.56	3.45	3.37	3.30	3.23	3.13	3.05	2.94	2.86	2.77	2.69	2.63	2.56	2.53	2.47	2.44	2.42	
21	4.32	3.47	3.07	2.84	2.68	2.57	2.49	2.42	2.37	2.32	2.28	2.25	2.20	2.15	2.09	2.05	2.00	1.96	1.93	1.89	1.87	1.84	1.82	1.81	21
	8.02	5.78	4.87	4.37	4.04	3.81	3.65	3.51	3.40	3.31	3.24	3.17	3.07	2.99	2.88	2.80	2.72	2.63	2.58	2.51	2.47	2.42	2.38	2.36	
22	4.30	3.44	3.05	2.82	2.66	2.55	2.47	2.40	2.35	2.30	2.26	2.23	2.18	2.13	2.07	2.03	1.98	1.93	1.91	1.87	1.84	1.81	1.80	1.78	22
	7.94	5.72	4.82	4.31	3.99	3.76	3.59	3.45	3.35	3.26	3.18	3.12	3.02	2.94	2.83	2.75	2.67	2.58	2.53	2.46	2.42	2.37	2.33	2.31	
23	4.28	3.42	3.03	2.80	2.64	2.53	2.45	2.38	2.32	2.28	2.24	2.20	2.14	2.10	2.04	2.00	1.96	1.91	1.88	1.84	1.82	1.79	1.77	1.76	23
	7.88	5.66	4.76	4.26	3.94	3.71	3.54	3.41	3.30	3.21	3.14	3.07	2.97	2.89	2.78	2.70	2.62	2.53	2.48	2.41	2.37	2.32	2.28	2.26	
24	4.26	3.40	3.01	2.78	2.62	2.51	2.43	2.36	2.30	2.26	2.22	2.18	2.13	2.09	2.02	1.98	1.94	1.89	1.86	1.82	1.80	1.76	1.74	1.73	24
	7.82	5.61	4.72	4.22	3.90	3.67	3.50	3.36	3.25	3.17	3.09	3.03	2.93	2.85	2.74	2.66	2.58	2.49	2.44	2.36	2.33	2.27	2.23	2.21	
25	4.24	3.38	2.99	2.76	2.60	2.49	2.41	2.34	2.28	2.24	2.20	2.16	2.11	2.06	2.00	1.96	1.92	1.87	1.84	1.80	1.77	1.74	1.72	1.71	25
	7.77	5.57	4.68	4.18	3.86	3.63	3.46	3.32	3.21	3.13	3.05	2.99	2.89	2.81	2.70	2.62	2.54	2.45	2.40	2.32	2.29	2.23	2.19	2.17	
26	4.22	3.37	2.98	2.74	2.59	2.47	2.39	2.32	2.27	2.22	2.18	2.15	2.10	2.05	1.99	1.95	1.90	1.85	1.82	1.78	1.76	1.72	1.70	1.69	26
	7.72	5.53	4.64	4.14	3.82	3.59	3.42	3.29	3.17	3.09	3.02	2.96	2.86	2.77	2.66	2.58	2.50	2.41	2.36	2.28	2.25	2.19	2.15	2.13	

Phụ lục VI: BẢNG PHÂN BỐ F (tiếp theo)

f_2	v_1 (Độ tự do của tử số)																								f_2
	1	2	3	4	5	6	7	8	9	10	11	12	14	16	20	24	30	40	50	75	100	200	500	∞	
27	4.21	3.35	2.96	2.73	2.57	2.46	2.37	2.30	2.25	2.20	2.16	2.13	2.08	2.03	1.97	1.93	1.88	1.84	1.80	1.76	1.74	1.71	1.68	1.67	27
	7.68	5.49	4.60	4.11	3.79	3.56	3.39	3.26	3.14	3.06	2.98	2.93	2.83	2.74	2.63	2.55	2.47	2.38	2.33	2.25	2.21	2.16	2.12	2.10	
28	4.20	3.34	2.95	2.71	2.56	2.44	2.36	2.29	2.24	2.19	2.15	2.12	2.06	2.02	1.96	1.91	1.87	1.81	1.78	1.75	1.72	1.69	1.67	1.65	28
	7.64	5.45	4.57	4.07	3.76	3.53	3.36	3.23	3.11	3.03	2.95	2.90	2.80	2.71	2.60	2.52	2.44	2.35	2.30	2.22	2.18	2.13	2.09	2.06	
29	4.18	3.33	2.93	2.70	2.54	2.43	2.35	2.28	2.22	2.18	2.14	2.10	2.05	2.00	1.94	1.90	1.85	1.80	1.77	1.73	1.71	1.68	1.65	1.64	29
	7.60	5.42	4.54	4.04	3.73	3.50	3.33	3.20	3.08	3.00	2.92	2.87	2.77	2.68	2.57	2.49	2.41	2.32	2.27	2.19	2.15	2.10	2.06	2.03	
30	4.17	3.32	2.92	2.69	2.53	2.42	2.34	2.27	2.21	2.16	2.12	2.09	2.04	1.99	1.93	1.89	1.84	1.79	1.76	1.72	1.69	1.66	1.64	1.62	30
	7.56	5.39	4.51	4.02	3.70	3.47	3.30	3.17	3.06	2.98	2.90	2.84	2.74	2.66	2.55	2.47	2.38	2.29	2.24	2.16	2.13	2.07	2.03	2.01	
32	4.15	3.30	2.90	2.67	2.51	2.40	2.32	2.25	2.19	2.14	2.10	2.07	2.02	1.97	1.91	1.86	1.82	1.76	1.74	1.69	1.67	1.64	1.61	1.59	32
	7.50	5.34	4.46	3.97	3.66	3.42	3.25	3.12	3.01	2.94	2.86	2.80	2.70	2.62	2.51	2.42	2.34	2.25	2.20	2.12	2.08	2.02	1.98	1.96	
34	4.13	3.28	2.88	2.65	2.49	2.38	2.30	2.23	2.17	2.12	2.08	2.05	2.00	1.95	1.89	1.84	1.80	1.74	1.71	1.67	1.64	1.61	1.59	1.57	34
	7.44	5.29	4.42	3.93	3.61	3.38	3.21	3.08	2.97	2.89	2.82	2.76	2.66	2.58	2.47	2.38	2.30	2.21	2.15	2.08	2.04	1.98	1.94	1.91	
36	4.11	3.26	2.86	2.63	2.48	2.36	2.28	2.21	2.15	2.10	2.06	2.03	1.98	1.93	1.87	1.82	1.78	1.72	1.69	1.65	1.62	1.59	1.56	1.55	36
	7.39	5.25	4.38	3.89	3.58	3.35	3.18	3.04	2.94	2.86	2.78	2.72	2.62	2.54	2.43	2.35	2.26	2.17	2.12	2.04	2.00	1.94	1.90	1.87	
38	4.10	3.25	2.85	2.62	2.46	2.35	2.26	2.19	2.14	2.09	2.05	2.02	1.96	1.92	1.85	1.80	1.76	1.71	1.67	1.63	1.60	1.57	1.54	1.53	38
	7.35	5.21	4.34	3.86	3.54	3.32	3.15	3.02	2.91	2.82	2.75	2.69	2.59	2.51	2.40	2.32	2.22	2.14	2.08	2.00	1.97	1.90	1.86	1.84	
40	4.08	3.23	2.84	2.61	2.45	2.34	2.25	2.18	2.12	2.07	2.04	2.00	1.95	1.90	1.84	1.79	1.74	1.69	1.66	1.61	1.59	1.55	1.53	1.51	40
	7.31	5.18	4.31	3.83	3.51	3.29	3.12	2.99	2.88	2.80	2.73	2.66	2.56	2.49	2.37	2.29	2.20	2.11	2.05	1.97	1.94	1.88	1.84	1.81	
42	4.07	3.22	2.83	2.59	2.44	2.32	2.24	2.17	2.11	2.06	2.02	1.99	1.94	1.89	1.82	1.78	1.73	1.68	1.64	1.60	1.57	1.54	1.51	1.49	42
	7.27	5.15	4.29	3.80	3.49	3.26	3.10	2.96	2.86	2.77	2.70	2.64	2.54	2.46	2.35	2.26	2.17	2.08	2.02	1.94	1.91	1.85	1.80	1.78	
44	4.06	3.21	2.82	2.58	2.43	2.31	2.23	2.16	2.10	2.05	2.01	1.98	1.92	1.88	1.81	1.76	1.72	1.66	1.63	1.58	1.56	1.52	1.50	1.48	44
	7.24	5.12	4.26	3.78	3.46	3.24	3.07	2.94	2.84	2.75	2.68	2.62	2.52	2.44	2.32	2.24	2.15	2.06	2.00	1.92	1.88	1.82	1.78	1.75	
46	4.05	3.20	2.81	2.57	2.42	2.30	2.22	2.14	2.09	2.04	2.00	1.97	1.91	1.87	1.80	1.75	1.71	1.65	1.62	1.57	1.54	1.51	1.48	1.46	46
	7.21	5.10	4.24	3.76	3.44	3.22	3.05	2.92	2.82	2.73	2.66	2.60	2.50	2.42	2.30	2.22	2.13	2.04	1.98	1.90	1.86	1.80	1.76	1.72	
48	4.04	3.19	2.80	2.56	2.41	2.30	2.21	2.14	2.08	2.03	1.99	1.96	1.90	1.86	1.79	1.74	1.70	1.64	1.61	1.56	1.53	1.50	1.47	1.45	48
	7.19	5.08	4.22	3.74	3.42	3.20	3.04	2.90	2.80	2.71	2.64	2.58	2.48	2.40	2.28	2.20	2.11	2.02	1.96	1.88	1.84	1.78	1.73	1.70	

v_2 (Độ tự do của mẫu số)

Phụ lục VI: BẢNG PHÂN BỐ F (5% in lợt và 1% in đậm)

df_2 (Độ tự do của mẫu số)	df_1 (Độ tự do của tử số)																								df_2
	1	2	3	4	5	6	7	8	9	10	11	12	14	16	20	24	30	40	50	75	100	200	500	∞	
50	4.03	3.18	2.79	2.56	2.40	2.29	2.20	2.13	2.07	2.02	1.98	1.95	1.90	1.85	1.78	1.74	1.69	1.63	1.60	1.55	1.52	1.48	1.46	1.44	50
	7.17	**5.06**	**4.20**	**3.72**	**3.41**	**3.18**	**3.02**	**2.88**	**2.78**	**2.70**	**2.62**	**2.56**	**2.46**	**2.39**	**2.26**	**2.18**	**2.10**	**2.00**	**1.94**	**1.86**	**1.82**	**1.76**	**1.71**	**1.68**	
55	4.02	3.17	2.78	2.54	2.38	2.27	2.18	2.11	2.05	2.00	1.97	1.93	1.88	1.83	1.76	1.72	1.67	1.61	1.58	1.52	1.50	1.46	1.43	1.41	55
	7.12	**5.01**	**4.16**	**3.68**	**3.37**	**3.15**	**2.98**	**2.85**	**2.75**	**2.66**	**2.59**	**2.53**	**2.43**	**2.35**	**2.23**	**2.15**	**2.06**	**1.96**	**1.90**	**1.82**	**1.78**	**1.71**	**1.66**	**1.64**	
60	4.00	3.15	2.76	2.52	2.37	2.25	2.17	2.10	2.04	1.99	1.95	1.92	1.86	1.81	1.75	1.70	1.65	1.59	1.56	1.50	1.48	1.44	1.41	1.39	60
	7.08	**4.98**	**4.13**	**3.65**	**3.34**	**3.12**	**2.95**	**2.82**	**2.72**	**2.63**	**2.56**	**2.50**	**2.40**	**2.32**	**2.20**	**2.12**	**2.03**	**1.93**	**1.87**	**1.79**	**1.74**	**1.68**	**1.63**	**1.60**	
65	3.99	3.14	2.75	2.51	2.36	2.24	2.15	2.08	2.02	1.98	1.94	1.90	1.85	1.80	1.73	1.68	1.63	1.57	1.54	1.49	1.46	1.42	1.39	1.37	65
	7.04	**4.95**	**4.10**	**3.82**	**3.31**	**3.09**	**2.93**	**2.79**	**2.70**	**2.61**	**2.54**	**2.47**	**2.37**	**2.30**	**2.18**	**2.09**	**2.00**	**1.90**	**1.84**	**1.76**	**1.71**	**1.64**	**1.60**	**1.56**	
70	3.98	3.13	2.74	2.50	2.35	2.23	2.14	2.07	2.01	1.97	1.93	1.89	1.84	1.79	1.72	1.67	1.62	1.56	1.53	1.47	1.45	1.40	1.37	1.36	70
	7.01	**4.92**	**4.08**	**3.60**	**3.29**	**3.07**	**2.91**	**2.77**	**2.67**	**2.59**	**2.51**	**2.45**	**2.35**	**2.28**	**2.15**	**2.07**	**1.98**	**1.88**	**1.82**	**1.74**	**1.69**	**1.62**	**1.56**	**1.53**	
80	3.96	3.11	2.72	2.48	2.33	2.21	2.12	2.05	1.99	1.95	1.91	1.88	1.82	1.77	1.70	1.65	1.60	1.54	1.51	1.45	1.42	1.38	1.35	1.32	80
	6.96	**4.88**	**4.04**	**3.56**	**3.25**	**3.04**	**2.87**	**2.74**	**2.64**	**2.55**	**2.48**	**2.41**	**2.32**	**2.24**	**2.11**	**2.03**	**1.94**	**1.84**	**1.78**	**1.70**	**1.65**	**1.57**	**1.52**	**1.49**	
100	3.94	3.09	2.70	2.46	2.30	2.19	2.10	2.03	1.97	1.92	1.88	1.85	1.79	1.75	1.68	1.63	1.57	1.51	1.48	1.42	1.39	1.34	1.30	1.28	100
	6.90	**4.82**	**3.98**	**3.51**	**3.20**	**2.99**	**2.82**	**2.69**	**2.59**	**2.51**	**2.43**	**2.36**	**2.26**	**2.19**	**2.06**	**1.98**	**1.89**	**1.79**	**1.73**	**1.64**	**1.59**	**1.51**	**1.46**	**1.43**	
125	3.92	3.07	2.68	2.44	2.29	2.17	2.08	2.01	1.95	1.90	1.86	1.83	1.77	1.72	1.65	1.60	1.55	1.49	1.45	1.39	1.36	1.31	1.27	1.25	125
	6.84	**4.78**	**3.94**	**3.47**	**3.17**	**2.95**	**2.79**	**2.65**	**2.56**	**2.47**	**2.40**	**2.33**	**2.23**	**2.15**	**2.03**	**1.94**	**1.85**	**1.75**	**1.68**	**1.59**	**1.54**	**1.43**	**1.40**	**1.37**	
150	3.91	3.06	2.67	2.43	2.27	2.16	2.07	2.00	1.94	1.89	1.85	1.82	1.76	1.71	1.64	1.59	1.54	1.47	1.44	1.37	1.34	1.29	1.25	1.22	150
	6.81	**4.75**	**3.91**	**3.44**	**3.14**	**2.92**	**2.76**	**2.62**	**2.53**	**2.44**	**2.37**	**2.30**	**2.20**	**2.12**	**2.00**	**1.91**	**1.83**	**1.72**	**1.66**	**1.56**	**1.51**	**1.43**	**1.37**	**1.33**	
200	3.89	3.04	2.65	2.41	2.26	2.14	2.05	1.98	1.92	1.87	1.83	1.80	1.74	1.69	1.62	1.57	1.52	1.45	1.42	1.35	1.32	1.26	1.22	1.19	200
	6.76	**4.71**	**3.88**	**3.41**	**3.11**	**2.90**	**2.73**	**2.60**	**2.50**	**2.41**	**2.34**	**2.28**	**2.17**	**2.09**	**1.97**	**1.88**	**1.79**	**1.69**	**1.62**	**1.53**	**1.48**	**1.39**	**1.33**	**1.28**	
400	3.86	3.02	2.62	2.39	2.23	2.12	2.03	1.96	1.90	1.85	1.81	1.78	1.72	1.67	1.60	1.54	1.49	1.42	1.38	1.32	1.28	1.22	1.16	1.13	400
	6.70	**4.66**	**3.83**	**3.36**	**3.06**	**2.85**	**2.69**	**2.55**	**2.46**	**2.37**	**2.29**	**2.23**	**2.12**	**2.04**	**1.92**	**1.84**	**1.74**	**1.64**	**1.57**	**1.47**	**1.42**	**1.32**	**1.24**	**1.19**	
1000	3.85	3.00	2.61	2.38	2.22	2.10	2.02	1.95	1.89	1.84	1.80	1.76	1.70	1.65	1.58	1.53	1.47	1.41	1.36	1.30	1.26	1.19	1.13	1.08	1000
	6.66	**4.62**	**3.80**	**3.34**	**3.04**	**2.82**	**2.66**	**2.53**	**2.43**	**2.34**	**2.26**	**2.20**	**2.09**	**2.01**	**1.89**	**1.81**	**1.71**	**1.61**	**1.54**	**1.44**	**1.38**	**1.28**	**1.19**	**1.11**	
∞	3.84	2.99	2.60	2.37	2.21	2.09	2.01	1.94	1.88	1.83	1.79	1.75	1.69	1.64	1.57	1.52	1.46	1.40	1.35	1.28	1.24	1.17	1.11	1.00	∞
	6.64	**4.60**	**3.78**	**3.32**	**3.02**	**2.80**	**2.64**	**2.51**	**2.41**	**2.32**	**2.24**	**2.18**	**2.07**	**1.99**	**1.87**	**1.79**	**1.69**	**1.59**	**1.52**	**1.41**	**1.36**	**1.25**	**1.15**	**1.00**	

Phụ lục VII : BẢNG SỐ NGẪU NHIÊN

14620	95430	12951	81953	17629	83603	09137	26453	02148	30742
09724	85125	48477	42783	70473	52491	66875	93650	91487	37190
56919	17803	96781	85069	61594	85437	92086	53045	31847	36207
97310	78209	51263	52396	82681	82611	70858	78195	47615	23721
07686	28040	26939	64531	70570	98412	74070	83468	18295	32586
25950	86189	69374	37904	06759	70799	59249	63461	75108	45703
82973	16405	81497	20863	94072	83615	09701	47920	46857	31924
60819	27364	69081	72635	49180	72537	46950	81736	53290	81736
59041	38476	03615	84093	49731	62748	39206	47315	84697	30863
74208	69516	79530	47649	53046	95420	41857	69420	79762	01935
39412	03642	87497	29735	14308	46309	28493	75091	82753	15040
48480	60075	11804	24956	72182	59649	16284	83538	53920	47192
95318	28749	49512	35408	21814	07564	70949	50969	15395	26081
72094	16385	90185	72635	86259	38352	94710	36853	94969	38405
63158	49753	84279	56496	30618	23973	25354	25237	48544	20405
19082	73645	09182	73649	56823	95208	49635	01420	46768	45362
15232	84146	87729	65584	83641	19468	34739	57052	43056	29950
94252	77489	62434	20965	20247	03994	25989	19609	74372	74151
72020	18895	84948	53072	74573	19520	92764	85397	52095	18079
48392	06359	47040	05695	79799	05342	54212	21539	48207	95920
37960	77387	35495	48192	84518	30210	23805	27837	24953	42610
09394	59842	39573	51630	78548	06461	06566	21752	78967	45692
34800	28055	91570	99154	39603	76846	77183	50369	16501	68867
36435	75946	85712	06293	86621	97764	53126	37396	57039	06096
28187	31824	52265	80494	66428	15703	05792	53376	54205	91590
13838	79940	97007	67511	87939	68417	21786	09822	67510	23817
72201	08423	41489	15498	94911	79392	65362	19672	93682	84190
63435	45192	62020	47358	32286	41659	31842	47269	70904	62972
59038	96983	49218	57179	08062	25074	06374	96484	59159	23749
62367	46627	58317	76928	50274	28705	45060	50903	66578	41465
71254	81686	85861	63973	96086	89681	50212	92829	27698	62284
07896	62924	35682	42820	43646	37385	37236	16496	51396	77975
71433	64331	58437	03542	76797	50437	13576	72876	02323	95237
54614	19092	83860	11351	32533	56032	42009	49745	14651	80128
30176	71248	37983	06073	89096	43498	95782	70452	90804	12042
79072	87795	23294	61602	62921	38385	69546	47104	72917	66273
75014	96754	67151	82741	24283	64276	78438	70757	40749	85183
37390	75846	74579	94606	54959	35310	31249	15101	95390	73432
24524	32751	28350	43090	79672	94672	07091	42920	46046	38083
26316	20378	16474	62438	42496	35191	40368	30074	93436	29425
61085	96937	02520	86801	30980	58479	34924	25101	87373	61560
45836	41086	41283	97460	51798	29852	47271	42480	94156	49341
92103	19679	16921	65924	12521	31724	60336	01968	15971	07963
10317	82592	65205	12528	24367	15817	12479	52021	02350	76394
39764	21251	41749	43789	70565	35496	87172	76830	41843	83489
83594	95692	52910	23202	93736	10817	53164	10724	27035	67562
08087	01753	01787	51631	74978	79608	01242	07525	72656	80854
57819	39689	32509	87540	38150	47872	14614	18427	06725	69326
96957	81060	28587	60905	67404	80450	21082	16074	61437	24961
48426	43513	82950	79838	45149	07143	73967	23723	06909	75375

Phụ lục VIII : CÁC BỐ TRÍ HÌNH VUÔNG LATIN MẪU

3 x 3

A	B	C
B	C	A
C	A	B

4 x 4

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

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

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

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

5 x 5

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

6 x 6

A	B	C	D	E	F
B	F	D	C	A	E
C	D	E	F	B	A
D	A	F	E	C	B
E	C	A	B	F	D
F	E	B	A	D	C

7 x 7

A	B	C	D	E	F	G
B	C	D	E	F	G	A
C	D	E	F	G	A	B
D	E	F	G	A	B	C
E	F	G	A	B	C	D
F	G	A	B	C	D	E
G	A	B	C	D	E	F

8 x 8

A	B	C	D	E	F	G	H
B	C	D	E	F	G	H	A
C	D	E	F	G	H	A	B
D	E	F	G	H	A	B	C
E	F	G	H	A	B	C	D
F	G	H	A	B	C	D	E
G	H	A	B	C	D	E	F
H	A	B	C	D	E	F	G

9 x 9

A	B	C	D	E	F	G	H	I
B	C	D	E	F	G	H	I	A
C	D	E	F	G	H	I	A	B
D	E	F	G	H	I	A	B	C
E	F	G	H	I	A	B	C	D
F	G	H	I	A	B	C	D	E
G	H	I	A	B	C	D	E	F
H	I	A	B	C	D	E	F	G
I	A	B	C	D	E	F	G	H

10 x 10

A	B	C	D	E	F	G	H	I	J
B	C	D	E	F	G	H	I	J	A
C	D	E	F	G	H	I	J	A	B
D	E	F	G	H	I	J	A	B	C
E	F	G	H	I	J	A	B	C	D
F	G	H	I	J	A	B	C	D	E
G	H	I	J	A	B	C	D	E	F
H	I	J	A	B	C	D	E	F	G
I	J	A	B	C	D	E	F	G	H
J	A	B	C	D	E	F	G	H	I

11 x 11

A	B	C	D	E	F	G	H	I	J	K
B	C	D	E	F	G	H	I	J	K	A
C	D	E	F	G	H	I	J	K	A	B
D	E	F	G	H	I	J	K	A	B	C
E	F	G	H	I	J	K	A	B	C	D
F	G	H	I	J	K	A	B	C	D	E
G	H	I	J	K	A	B	C	D	E	F
H	I	J	K	A	B	C	D	E	F	G
I	J	K	A	B	C	D	E	F	G	H
J	K	A	B	C	D	E	F	G	H	I
K	A	B	C	D	E	F	G	H	I	J

Phụ lục IX: BẢNG DUNNET

($\alpha = 0,05$)

	Giá trị của d_t								
	t = Số nghiệm thức so với đối chứng								
df	1	2	3	4	5	6	7	8	9
5	2.02	2.44	2.68	2.85	2.98	3.08	3.16	3.24	3.30
6	1.94	2.34	2.56	2.71	2.83	2.92	3.00	3.07	3.12
7	1.89	2.27	2.48	2.62	2.73	2.82	2.89	2.05	3.01
8	1.86	2.22	2.42	2.55	2.66	2.74	2.81	2.87	2.92
9	1.83	2.18	2.37	2.50	2.60	2.68	2.75	2.81	2.86
10	1.81	2.15	2.34	2.47	2.56	2.64	2.70	2.76	2.81
11	1.80	2.13	2.31	2.44	2.53	2.60	2.67	2.72	2.77
12	1.78	2.11	2.29	2.41	2.50	2.58	2.64	2.69	2.74
13	1.77	2.09	2.27	2.39	2.48	2.55	2.61	2.66	2.71
14	1.76	2.08	2.25	2.37	2.46	2.53	2.59	2.64	2.69
15	1.75	2.07	2.24	2.36	2.44	2.51	2.57	2.62	2.67
16	1.75	2.06	2.23	2.34	2.43	2.50	2.56	2.61	2.65
17	1.74	2.05	2.22	2.33	2.42	2.49	2.54	2.59	2.64
18	1.73	2.04	2.21	2.32	2.41	2.48	2.53	2.58	2.62
19	1.73	2.03	2.20	2.31	2.40	2.47	2.52	2.57	2.61
20	1.72	2.03	2.19	2.30	2.39	2.46	2.51	2.56	2.60
24	1.71	2.01	2.17	2.28	2.36	2.43	2.48	2.53	2.57
30	1.70	1.99	2.15	2.25	2.33	2.40	2.45	2.50	2.54
40	1.68	1.97	2.13	2.23	2.31	2.37	2.42	2.47	2.51
60	1.67	1.95	2.10	2.21	2.28	2.35	2.39	2.44	2.48
120	1.66	1.93	2.08	2.18	2.26	2.32	2.37	2.41	2.45
∞	1.64	1.92	2.06	2.16	2.23	2.29	2.34	2.38	2.42

($\alpha = 0,01$)

	Giá trị của d_t								
	t = Số nghiệm thức so với đối chứng								
df	1	2	3	4	5	6	7	8	9
5	3.37	3.90	4.21	4.43	4.60	4.73	4.85	4.94	5.03
6	3.14	3.61	3.88	4.07	4.21	4.33	4.43	4.51	4.59
7	3.00	3.42	3.66	3.83	3.96	4.07	4.15	4.23	4.30
8	2.90	3.29	3.51	3.67	3.79	3.88	3.96	4.03	4.09
9	2.82	3.19	3.40	3.55	3.66	3.75	3.82	3.89	3.94
10	2.76	3.11	3.31	3.45	3.56	3.64	3.71	3.78	3.83
11	2.72	3.06	3.25	3.38	3.48	3.56	3.63	3.69	3.74
12	2.68	3.01	3.19	3.32	3.42	3.50	3.56	3.62	3.67
13	2.65	2.97	3.15	3.27	3.37	3.44	3.51	3.56	3.61
14	2.62	2.94	3.11	3.23	3.32	3.40	3.46	3.51	3.56
15	2.60	2.91	3.08	3.20	3.29	3.36	3.42	3.47	3.52
16	2.58	2.88	3.05	3.17	3.26	3.33	3.39	3.44	3.48
17	2.57	2.86	3.03	3.14	3.23	3.30	3.36	3.41	3.45
18	2.55	2.84	3.01	3.12	3.21	3.27	3.33	3.38	3.42
19	2.54	2.83	2.99	3.10	3.18	3.25	3.31	3.36	3.40
20	2.53	2.81	2.97	3.08	3.17	3.23	3.29	3.34	3.38
24	2.49	2.77	2.92	3.03	3.11	3.17	3.22	3.27	3.31
30	2.46	2.72	2.87	2.97	3.05	3.11	3.16	3.21	3.24
40	2.42	2.68	2.82	2.92	2.99	3.05	3.10	3.14	3.18
60	2.39	2.64	2.78	2.87	2.94	3.00	3.04	3.08	3.12
120	2.36	2.60	2.73	2.82	2.89	2.94	2.99	3.03	3.06
∞	2.33	2.56	2.68	2.77	2.84	2.89	2.93	2.97	3.00

Phụ lục X: BẢNG DUNCAN Ở MỨC Ý NGHĨA 5% VÀ 1%

df sai số	Mức α	p = Số trung bình chứa trong khoảng đang được kiểm định													
		2	3	4	5	6	7	8	9	10	12	14	16	18	20
1	.05	18.0	18.0	18.0	18.0	18.0	18.0	18.0	18.0	18.0	18.0	18.0	18.0	18.0	18.0
	.01	90.0	90.0	90.0	90.0	90.0	90.0	90.0	90.0	90.0	90.0	90.0	90.0	90.0	90.0
2	.05	6.09	6.09	6.09	6.09	6.09	6.09	6.09	6.09	6.09	6.09	6.09	6.09	6.09	6.09
	.01	14.0	14.0	14.0	14.0	14.0	14.0	14.0	14.0	14.0	14.0	14.0	14.0	14.0	14.0
3	.05	4.50	4.50	4.50	4.50	4.50	4.50	4.50	4.50	4.50	4.50	4.50	4.50	4.50	4.50
	.01	8.26	8.5	8.6	8.7	8.8	8.9	8.9	9.0	9.0	9.0	9.1	9.2	9.3	9.3
4	.05	3.93	4.01	4.02	4.02	4.02	4.02	4.02	4.02	4.02	4.02	4.02	4.02	4.02	4.02
	.01	6.51	6.8	6.9	7.0	7.1	7.1	7.2	7.2	7.3	7.3	7.4	7.4	7.5	7.5
5	.05	3.64	3.74	3.79	3.83	3.83	3.83	3.83	3.83	3.83	3.83	3.83	3.83	3.83	3.83
	.01	5.70	5.96	6.11	6.18	6.26	6.33	6.40	6.44	6.5	6.6	6.6	6.7	6.7	6.8
6	.05	3.46	3.58	3.64	3.68	3.68	3.68	3.68	3.68	3.68	3.68	3.68	3.68	3.68	3.68
	.01	5.24	5.51	5.65	5.73	5.81	5.88	5.95	6.00	6.0	6.1	6.2	6.2	6.3	6.3
7	.05	3.35	3.47	3.54	3.58	3.60	3.61	3.61	3.61	3.61	3.61	3.61	3.61	3.61	3.61
	.01	4.95	5.22	5.37	5.45	5.53	5.61	5.69	5.73	5.8	5.8	5.9	5.9	6.0	6.0
8	.05	3.26	3.39	3.47	3.52	3.55	3.56	3.56	3.56	3.56	3.56	3.56	3.56	3.56	3.56
	.01	4.74	5.00	5.14	5.23	5.32	5.40	5.47	5.51	5.5	5.6	5.7	5.7	5.8	5.8
9	.05	3.20	3.34	3.41	3.47	3.50	3.52	3.52	3.52	3.52	3.52	3.52	3.52	3.52	3.52
	.01	4.60	4.86	4.99	5.08	5.17	5.25	5.32	5.36	5.4	5.5	5.5	5.6	5.7	5.7
10	.05	3.15	3.30	3.37	3.43	3.46	3.47	3.47	3.47	3.47	3.47	3.47	3.47	3.47	3.48
	.01	4.48	4.73	4.88	4.96	5.06	5.13	5.20	5.24	5.28	5.36	5.42	5.48	5.54	5.55
11	.05	3.11	3.27	3.35	3.39	3.43	3.44	3.45	3.46	4.46	3.46	3.46	3.46	3.47	3.48
	.01	4.39	4.63	4.77	4.86	4.94	5.01	5.06	5.12	5.15	5.24	5.28	5.34	5.38	5.39
12	.05	3.08	3.23	3.33	3.36	3.40	3.42	3.44	3.44	3.46	3.46	3.46	3.46	3.47	3.48
	.01	4.32	4.55	4.68	4.76	4.81	4.92	4.96	5.02	5.07	5.13	5.17	5.22	5.24	5.26
13	.05	3.06	3.21	3.30	3.35	3.38	3.41	3.42	3.44	3.45	3.45	3.46	3.46	3.47	3.47
	.01	4.26	4.48	4.62	4.69	4.74	4.84	4.88	4.94	4.98	5.04	5.08	5.13	5.14	5.15
14	.05	3.03	3.18	3.27	3.33	3.37	3.39	3.41	3.42	3.44	3.45	3.46	3.46	3.47	3.47
	.01	4.21	4.42	4.55	4.63	4.70	4.78	4.83	3.87	4.91	4.96	5.00	5.04	5.06	5.07
15	.05	3.01	3.16	3.25	3.31	3.36	3.38	3.40	3.42	3.43	3.44	3.45	3.46	3.47	3.47
	.01	4.17	4.37	4.50	4.58	4.64	4.72	4.77	4.81	4.84	4.90	4.94	4.97	4.99	5.00

Phụ lục X: BẢNG DUNCAN Ở MỨC Ý NGHĨA 5% VÀ 1% (Tiếp theo)

df sai số	Mức α	p = Số trung bình chứa trong khoảng đang kiểm định													
		2	3	4	5	6	7	8	9	10	12	14	16	18	20
16	.05	3.00	3.15	3.23	3.30	3.34	3.37	3.39	3.41	3.43	3.44	3.45	3.46	3.47	3.47
	.01	4.13	4.34	4.45	4.54	4.60	4.67	4.72	4.76	4.79	4.84	4.88	4.91	4.93	4.94
17	.05	2.98	3.13	3.22	3.28	3.33	3.36	3.38	3.40	3.42	3.44	3.45	3.46	3.47	3.47
	.01	4.10	4.30	4.41	4.50	4.56	4.63	4.68	4.72	4.75	4.80	4.83	4.86	4.88	4.89
18	.05	2.97	3.12	3.21	3.27	3.32	3.35	3.37	3.39	3.41	3.43	3.45	3.46	3.47	3.47
	.01	4.07	4.27	4.38	4.46	4.53	4.59	4.64	4.68	4.71	4.76	4.79	4.82	4.84	4.85
19	.05	2.96	3.11	3.19	3.26	3.31	3.35	3.37	3.39	3.41	3.43	3.44	3.46	3.47	3.47
	.01	4.05	4.24	4.35	4.43	4.50	4.56	4.61	4.64	4.67	4.72	4.76	4.79	4.81	4.82
20	.05	2.95	3.10	3.18	3.25	3.30	3.34	3.36	3.38	3.40	3.43	3.44	3.46	3.46	3.47
	.01	4.02	4.22	4.33	4.40	4.47	4.53	4.58	4.61	4.65	4.69	4.73	3.76	4.78	4.79
22	.05	2.93	3.08	3.17	3.24	3.29	3.32	3.35	3.37	3.39	3.42	3.44	3.45	3.46	3.47
	.01	3.99	4.17	4.28	4.36	4.42	4.48	4.53	4.57	4.60	4.65	4.68	4.71	4.74	4.75
24	.05	2.92	3.07	3.15	3.22	3.28	3.31	3.34	3.37	3.38	3.41	3.44	3.45	3.46	3.47
	.01	3.96	4.14	4.24	4.33	4.39	4.44	4.49	4.53	4.57	4.62	4.64	4.67	4.70	4.72
26	.05	2.91	3.06	3.14	3.21	3.27	3.30	3.34	3.36	3.38	3.41	3.43	3.45	3.46	3.47
	.01	3.93	4.11	4.21	4.30	4.36	4.41	4.46	4.50	4.53	4.58	4.62	4.65	4.67	4.69
28	.05	2.90	3.04	3.13	3.20	3.26	3.30	3.33	3.35	3.37	3.40	3.43	3.45	3.46	3.47
	.01	3.91	4.08	4.18	4.28	4.34	4.39	4.43	4.47	4.51	4.56	4.60	4.62	4.65	4.67
30	.05	2.89	3.04	3.12	3.20	3.25	3.29	3.32	3.35	3.37	3.40	3.43	3.44	3.46	3.47
	.01	3.89	4.06	4.16	4.22	4.32	4.36	4.41	4.45	4.48	4.54	4.58	4.61	4.63	4.65
40	.05	2.86	3.01	3.10	3.17	3.22	3.27	3.30	3.33	3.35	3.39	3.42	3.44	3.46	3.47
	.01	3.82	3.99	4.10	4.17	4.21	4.30	4.34	4.37	4.41	4.46	4.51	4.54	4.57	4.59
60	.05	2.83	2.98	3.08	3.14	3.20	3.24	3.28	3.31	3.33	3.37	3.40	3.43	3.45	3.47
	.01	3.76	3.92	4.03	4.12	4.17	4.23	4.27	4.31	4.34	4.39	4.44	4.47	4.50	4.53
100	.05	2.80	2.95	3.05	3.12	3.18	3.22	3.26	3.29	3.32	3.36	3.40	3.42	3.45	3.47
	.01	3.71	3.86	3.98	4.06	4.11	4.17	4.21	4.25	4.29	4.35	4.38	4.42	4.45	4.48
∞	.05	2.77	2.92	3.02	3.09	3.15	3.19	3.23	3.26	3.29	3.34	3.38	3.41	3.44	3.47
	.01	3.64	3.80	3.90	3.98	4.04	4.09	4.14	4.17	4.20	4.26	4.31	4.34	4.38	4.41

Phụ lục XI : CÁC GIÁ TRỊ CHUẨN CỦA KIỂM ĐỊNH NEWMAN VÀ KEULS ($\alpha = 0,05$)

k \ p	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
2	6,08	8,33	9,80	10,88	11,74	12,44	13,03	13,54	13,99	14,39	14,75	15,08	15,38	15,65	15,91	16,14	16,37	16,57	16,77
3	4,50	5,91	6,82	7,50	8,04	8,48	8,85	9,18	9,46	9,72	9,95	10,15	10,35	10,52	10,69	10,84	10,98	11,11	11,24
4	3,93	5,04	5,76	6,29	6,71	7,05	7,35	7,60	7,83	8,03	8,21	8,37	8,52	8,66	8,79	8,91	9,03	9,13	9,23
5	3,64	4,60	5,22	5,67	6,03	6,33	6,58	6,80	6,99	7,17	7,32	7,47	7,60	7,72	7,83	7,93	8,03	8,12	8,21
6	3,46	4,34	4,90	5,30	5,63	5,90	6,12	6,32	6,49	6,65	6,79	6,92	7,03	7,14	7,24	7,34	7,43	7,51	7,59
7	3,34	4,16	4,68	5,06	5,36	5,61	5,82	6,00	6,16	6,30	6,43	6,55	6,66	6,76	6,85	6,94	7,02	7,10	7,17
8	3,26	4,04	4,53	4,89	5,17	5,40	5,60	5,77	5,92	6,05	6,18	6,29	6,39	6,48	6,57	6,65	6,73	6,80	6,87
9	3,20	3,95	4,41	4,76	5,02	5,24	5,43	5,59	5,74	5,87	5,98	6,09	6,19	6,28	6,36	6,44	6,51	6,58	6,64
10	3,15	3,88	4,33	4,65	4,91	5,12	5,30	5,46	5,60	5,72	5,83	5,93	6,03	6,11	6,19	6,27	6,34	6,40	6,47
11	3,11	3,82	4,26	4,57	4,82	5,03	5,20	5,35	5,49	5,61	5,71	5,81	5,90	5,98	6,06	6,13	6,20	6,27	6,33
12	3,08	3,77	4,20	4,51	4,75	4,95	5,12	5,27	5,39	5,51	5,61	5,71	5,80	5,88	5,95	6,02	6,09	6,15	6,21
13	3,06	3,73	4,15	4,45	4,69	4,88	5,05	5,19	5,32	5,43	5,53	5,63	5,71	5,79	5,86	5,93	5,99	6,05	6,11
14	3,03	3,70	4,11	4,41	4,64	4,83	4,99	5,13	5,25	5,36	5,46	5,55	5,64	5,71	5,79	5,85	5,91	5,97	6,03
15	3,01	3,67	4,08	4,37	4,59	4,78	4,94	5,08	5,20	5,31	5,40	5,49	5,57	5,65	5,72	5,78	5,85	5,90	5,96
16	3,00	3,65	4,05	4,33	4,56	4,74	4,90	5,03	5,15	5,26	5,35	5,44	5,52	5,59	5,66	5,73	5,79	5,84	5,90
17	2,98	3,63	4,02	4,30	4,52	4,70	4,86	4,99	5,11	5,21	5,31	5,39	5,47	5,54	5,61	5,67	5,73	5,79	5,84
18	2,97	3,61	4,00	4,28	4,49	4,67	4,82	4,96	5,07	5,17	5,27	5,35	5,43	5,50	5,57	5,63	5,69	5,74	5,79
19	2,96	3,59	3,98	4,25	4,47	4,65	4,79	4,92	5,04	5,14	5,23	5,31	5,39	5,46	5,53	5,59	5,65	5,70	5,75
20	2,95	3,58	3,96	4,23	4,45	4,62	4,77	4,90	5,01	5,11	5,20	5,28	5,36	5,43	5,49	5,55	5,61	5,66	5,71
24	2,92	3,53	3,90	4,17	4,37	4,54	4,68	4,81	4,92	5,01	5,10	5,18	5,25	5,32	5,38	5,44	5,49	5,55	5,59
30	2,89	3,49	3,85	4,10	4,30	4,46	4,60	4,72	4,82	4,92	5,00	5,08	5,15	5,21	5,27	5,33	5,38	5,43	5,47
40	2,86	3,44	3,79	4,04	4,23	4,39	4,52	4,63	4,73	4,82	4,90	4,98	5,04	5,11	5,16	5,22	5,27	5,31	5,36
60	2,83	3,40	3,74	3,98	4,16	4,31	4,44	4,55	4,65	4,73	4,81	4,88	4,94	5,00	5,06	5,11	5,15	5,20	5,24
120	2,80	3,36	3,68	3,92	4,10	4,24	4,36	4,47	4,56	4,64	4,71	4,78	4,84	4,90	4,95	5,00	5,04	5,09	5,13
∞	2,77	3,31	3,63	3,86	4,03	4,17	4,29	4,39	4,47	4,55	4,62	4,68	4,74	4,80	4,85	4,89	4,93	4,97	5,01

Phụ lục XI : BẢNG CÁC GIÁ TRỊ CHUẨN CỦA KIỂM ĐỊNH NEWMAN VÀ KEULS ($\alpha = 0,01$)

k \ p	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
2	14,04	19,02	22,29	24,72	26,63	28,20	29,53	30,68	31,69	32,59	33,40	34,13	34,81	35,43	36,00	36,53	37,03	37,50	37,95
3	8,26	10,62	12,17	13,33	14,24	15,00	15,64	16,20	16,69	17,13	17,53	17,89	18,22	18,52	18,81	19,07	19,32	19,55	19,77
4	6,51	8,12	9,17	9,96	10,58	11,10	11,55	11,93	12,27	12,57	12,84	13,09	13,32	13,53	13,73	13,91	14,08	14,24	14,40
5	5,70	6,98	7,80	8,42	8,91	9,32	9,67	9,97	10,24	10,48	10,70	10,89	11,08	11,24	11,40	11,55	11,68	11,81	11,93
6	5,24	6,33	7,03	7,56	7,97	8,32	8,61	8,87	9,10	9,30	9,48	9,65	9,81	9,95	10,08	10,21	10,32	10,43	10,54
7	4,95	5,92	6,54	7,01	7,37	7,68	7,94	8,17	8,37	8,55	8,71	8,86	9,00	9,12	9,24	9,35	9,46	9,55	9,65
8	4,75	5,64	6,20	6,62	6,96	7,24	7,47	7,68	7,86	8,03	8,18	8,31	8,44	8,55	8,66	8,76	8,85	8,94	9,03
9	4,60	5,43	5,96	6,35	6,66	6,91	7,13	7,33	7,49	7,65	7,78	7,91	8,03	8,13	8,23	8,33	8,41	8,49	8,57
10	4,48	5,27	5,77	6,14	6,43	6,67	6,87	7,05	7,21	7,36	7,49	7,60	7,71	7,81	7,91	7,99	8,08	8,15	8,23
11	4,39	5,15	5,62	5,97	6,25	6,48	6,67	6,84	6,99	7,13	7,25	7,36	7,46	7,56	7,65	7,73	7,81	7,88	7,95
12	4,32	5,05	5,50	5,84	6,10	6,32	6,51	6,67	6,81	6,94	7,06	7,17	7,26	7,36	7,44	7,52	7,59	7,66	7,73
13	4,26	4,96	5,40	5,73	5,98	6,19	6,37	6,53	6,67	6,79	6,90	7,01	7,10	7,19	7,27	7,35	7,42	7,48	7,55
14	4,21	4,89	5,32	5,63	5,88	6,08	6,26	6,41	6,54	6,66	6,77	6,87	6,96	7,05	7,13	7,20	7,27	7,33	7,39
15	4,17	4,84	5,25	5,56	5,80	5,99	6,16	6,31	6,44	6,55	6,66	6,76	6,84	6,93	7,00	7,07	7,14	7,20	7,26
16	4,13	4,79	5,19	5,49	5,72	5,92	6,08	6,22	6,35	6,46	6,56	6,66	6,74	6,82	6,90	6,97	7,03	7,09	7,15
17	4,10	4,74	5,14	5,43	5,66	5,85	6,01	6,15	6,27	6,38	6,48	6,57	6,66	6,73	6,81	6,87	6,94	7,00	7,05
18	4,07	4,70	5,09	5,38	5,60	5,79	5,94	6,08	6,20	6,31	6,41	6,50	6,58	6,65	6,73	6,79	6,85	6,91	6,97
19	4,05	4,67	5,05	5,33	5,55	5,73	5,89	6,02	6,14	6,25	6,34	6,43	6,51	6,58	6,65	6,72	6,78	6,84	6,89
20	4,02	4,64	5,02	5,29	5,51	5,69	5,84	5,97	6,09	6,19	6,28	6,37	6,45	6,52	6,59	6,65	6,71	6,77	6,82
24	3,96	4,55	4,91	5,17	5,37	5,54	5,69	5,81	5,92	6,02	6,11	6,19	6,26	6,33	6,39	6,45	6,51	6,56	6,61
30	3,89	4,45	4,80	5,05	5,24	5,40	5,54	5,65	5,76	5,85	5,93	6,01	6,08	6,14	6,20	6,26	6,31	6,36	6,41
40	3,82	4,37	4,70	4,93	5,11	5,26	5,39	5,50	5,60	5,69	5,76	5,83	5,90	5,96	6,02	6,07	6,12	6,16	6,21
60	3,76	4,28	4,59	4,82	4,99	5,13	5,25	5,36	5,45	5,53	5,60	5,67	5,73	5,78	5,84	5,89	5,93	5,97	6,01
120	3,70	4,20	4,50	4,71	4,87	5,01	5,12	5,21	5,30	5,37	5,44	5,50	5,56	5,61	5,66	5,71	5,75	5,79	5,83
∞	3,64	4,12	4,40	4,60	4,76	4,88	4,99	5,08	5,16	5,23	5,29	5,35	5,40	5,45	5,49	5,54	5,57	5,61	5,65

Phụ lục XII: BẢNG HỆ SỐ TƯƠNG QUAN TUYẾN TÍNH (r) ở mức ý nghĩa 5% và 1%.

d.f.	5%	1%	d.f.	5%	1%
1	.997	1.000	26	.374	.478
2	.950	.990	27	.367	.470
3	.878	.959	28	.361	.463
4	.811	.917	29	.355	.456
5	.754	.874	30	.349	.449
6	.707	.834	32	.339	.437
7	.666	.798	34	.329	.424
8	.632	.765	36	.321	.413
9	.602	.735	38	.312	.403
10	.576	.708	40	.304	.393
11	.553	.684	45	.288	.372
12	.532	.661	50	.273	.354
13	.514	.641	55	.262	.340
14	.497	.623	60	.250	.325
15	.482	.606	70	.232	.302
16	.468	.590	80	.217	.283
17	.456	.575	90	.205	.267
18	.444	.561	100	.195	.254
19	.433	.549	125	.174	.228
20	.423	.537	150	.159	.208
21	.413	.526	175	.148	.194
22	.404	.515	200	.138	.181
23	.396	.505	300	.113	.148
24	.388	.496	400	.098	.128
25	.381	.487	500	.088	.115

df = n -2 với n là số cặp biến

Phụ lục XIII: BẢNG CÁC GIÁ TRỊ TƯƠNG ỨNG CỦA r[1] VÀ z[1]

z	r	z	r	z	r	z	r
0.00	0.000	0.70	0.607	1.40	0.885	2.10	0.970
0.20	0.020	0.72	0.617	1.42	0.890	2.12	0.972
0.24	0.040	0.74	0.629	1.44	0.894	2.14	0.973
0.06	0.060	0.76	0.641	1.46	0.898	2.16	0.974
0.08	0.080	0.80	0.653	1.48	0.902	2.18	0.975
0.10	0.100	0.82	0.664	1.50	0.905	2.20	0.976
0.12	0.119	0.84	0.675	1.52	0.909	2.22	0.977
0.14	0.139	0.86	0.686	1.54	0.912	2.24	0.978
0.16	0.159	0.88	0.696	1.56	0.915	2.26	0.978
0.18	0.178	0.90	0.706	1.58	0.919	2.28	0.979
0.20	0.197	0.92	0.716	1.60	0.922	2.30	0.980
0.22	0.216	0.94	0.726	1.62	0.925	2.32	0.981
0.24	0.236	0.96	0.735	1.64	0.928	2.34	0.982
0.26	0.254	0.98	0.744	1.66	0.930	2.36	0.982
0.28	0.273	1.00	0.753	1.68	0.933	2.38	0.983
0.30	0.291	1.02	0.762	1.70	0.935	2.40	0.984
0.32	0.310	1.04	0.770	1.72	0.938	2.42	0.984
0.34	0.327	1.06	0.778	1.74	0.940	2.44	0.985
0.36	0.345	1.08	0.786	1.76	0.942	2.46	0.986
0.38	0.363	1.10	0.793	1.78	0.945	2.48	0.986
0.40	0.380	1.12	0.800	1.80	0.947	2.50	0.987
0.42	0.397	1.14	0.808	1.82	0.949	2.52	0.987
0.44	0.414	1.16	0.814	1.84	0.951	2.54	0.988
0.46	0.430	1.18	0.821	1.86	0.953	2.56	0.988
0.48	0.446	1.20	0.828	1.88	0.954	2.58	0.989
0.50	0.462	1.22	0.834	1.90	0.956	2.62	0.989
0.52	0.478	1.24	0.840	1.92	0.958	2.66	0.990
0.54	0.493	1.26	0.846	1.94	0.960	2.70	0.991
0.56	0.508	1.28	0.851	1.96	0.961	2.74	0.992
0.58	0.523	1.30	0.856	1.98	0.963	2.78	0.992
0.60	0.537	1.32	0.862	2.00	0.964	2.82	0.993
0.62	0.551	1.34	0.867	2.02	0.965	2.86	0.993
0.64	0.565	1.36	0.872	2.04	0.967	2.90	0.994
0.66	0.578	1.38	0.876	2.06	0.968	2.94	0.994
0.68	0.592	1.40	0.881	2.08	0.969	2.98	0.995

[1] $r = (e^{2z} - 1)/(e^{2z} + 1)$ hoặc $z = ½ \ln(1 + r) / (1 - r)$

Trích từ STATISTICAL METHODS của William G. Cochran. 1946. 4th edition. Iowa State University Press. Ames. Iowa.

Phụ lục XIII: BẢNG ĐỔI SANG ARC SIN $\sqrt{\text{phần trăm}}$

[Chuyển số phần trăm nhị thức, đổi từ độ sang góc. Dấu (+) hoặc (-) theo sau các góc tận cùng bằng số 5 đã được làm tròn đến một số lẻ]

%	0	1	2	3	4	5	6	7	8	9
0.0	0	0.57	0.81	0.99	1.15-	1.28	1.40	1.52	1.62	1.72
0.1	1.81	1.90	1.99	2.07	2.14	2.22	2.29	2.36	2.43	2.50
0.2	2.56	2.63	2.69	2.75-	2.81	2.87	2.92	2.98	3.03	3.09
0.3	3.14	3.19	3.24	3.29	3.34	3.39	3.44	3.49	3.53	3.58
0.4	3.63	3.67	3.72	3.76	3.80	3.85-	3.89	3.93	3.97	4.01
0.5	4.05+	4.09	4.13	4.17	4.21	4.25+	4.29	4.33	4.37	4.40
0.6	4.44	4.48	4.52	4.55+	4.59	4.62	4.66	4.69	4.73	4.76
0.7	4.80	4.83	4.87	4.90	4.93	4.97	5.00	5.03	5.07	5.10
0.8	5.13	5.16	5.20	5.23	5.26	5.29	5.32	5.35+	5.38	5.41
0.9	5.44	5.47	5.50	5.53	5.56	5.59	5.62	5.65+	5.68	5.71
1	5.74	6.02	6.29	6.55-	6.80	7.04	7.27	7.49	7.71	7.92
2	8.13	8.33	8.53	8.72	8.91	9.10	9.28	9.46	9.63	9.81
3	9.98	10.14	10.31	10.47	10.63	10.78	10.94	11.09	11.24	11.39
4	11.54	11.68	11.83	11.97	12.11	12.25-	12.39	12.52	12.66	12.79
5	12.92	13.05+	13.18	13.31	13.44	13.56	13.69	13.81	13.94	14.06
6	14.18	14.30	14.42	14.54	14.65+	14.77	14.89	15.00	15.12	15.23
7	15.34	15.45+	15.56	15.68	15.79	15.89	16.00	16.11	16.22	16.32
8	16.43	16.54	16.64	16.74	16.85-	16.95+	17.05+	17.16	17.26	17.36
9	17.46	17.56	17.66	17.76	17.85+	17.95+	18.05-	18.15-	18.24	18.34
10	18.44	18.53	18.63	18.72	18.81	18.91	19.00	19.09	19.19	19.28
11	19.37	19.46	19.55+	19.64	19.73	19.82	19.91	20.00	20.09	20.18
12	20.27	20.36	20.44	20.53	20.62	20.70	20.79	20.88	20.96	21.05-
13	21.13	21.22	21.30	21.39	21.47	21.56	21.64	21.72	21.81	21.89
14	21.97	22.06	22.14	22.22	22.30	22.38	22.46	22.55-	22.63	22.71
15	22.79	22.87	22.95-	23.03	23.11	23.19	23.26	23.34	23.42	23.50
16	23.58	23.66	23.73	23.81	23.89	23.97	24.04	24.12	24.20	24.27
17	24.35+	24.43	24.50	24.58	24.65+	24.73	24.80	24.88	24.95+	25.03
18	25.10	25.18	25.25+	25.33	25.40	25.48	25.55-	25.62	25.70	25.77
19	25.84	25.92	25.99	26.06	26.13	26.21	26.28	26.35-	26.42	26.49
20	26.56	26.64	26.71	26.78	26.85+	26.92	26.99	27.06	27.13	27.20
21	27.28	27.35-	27.42	27.49	27.56	27.63	27.69	27.76	27.83	27.90
22	27.97	28.04	28.11	28.18	28.25-	28.32	28.38	28.45+	28.52	28.59
23	28.66	28.73	28.79	28.86	28.93	29.00	29.06	29.13	29.20	29.27
24	29.33	29.40	29.47	29.53	29.60	29.67	29.73	29.80	29.87	29.93
25	30.00	30.07	30.13	30.20	30.26	30.33	30.40	30.46	30.53	30.59
26	30.66	30.72	30.79	30.85+	30.92	30.98	31.05-	31.11	31.18	31.24
27	31.31	31.37	31.44	31.50	31.56	31.63	31.69	31.76	31.82	31.88
28	31.95-	32.01	32.08	32.14	32.20	32.27	32.33	32.39	32.46	32.52
29	32.58	32.65-	32.71	32.77	32.83	32.90	32.96	33.02	33.09	33.15-
30	33.21	33.27	33.34	33.40	33.46	33.52	33.58	33.65-	33.71	33.77
31	33.83	33.89	33.96	34.02	34.08	34.14	34.20	34.27	34.33	34.39
32	34.45-	34.51	34.57	34.63	34.70	34.76	34.82	34.88	34.94	35.00
33	35.06	35.12	35.18	35.24	35.30	35.37	35.43	35.49	35.55	35.61
34	35.67	35.73	35.79	35.85	35.91	35.97	36.03	36.09	36.15+	36.21
35	36.27	36.33	36.39	36.45+	36.51	36.57	36.63	36.69	36.75+	36.81
36	36.87	36.93	36.99	37.05-	37.11	37.17	37.23	37.29	37.35-	37.41
37	37.47	37.52	37.58	37.64	37.70	37.76	37.82	37.88	37.94	38.00
38	38.06	38.12	38.17	38.23	38.29	38.35+	38.41	38.47	38.53	38.59
39	38.65-	38.70	38.76	38.82	38.88	38.94	39.00	39.06	39.11	39.17
40	39.23	39.29	39.35-	39.41	39.47	39.52	39.58	39.64	39.70	39.76
41	39.82	39.87	39.93	39.99	40.05-	40.11	40.16	40.22	40.28	40.34
42	40.40	40.46	40.51	40.57	40.63	40.69	40.74	40.80	40.86	40.92
43	40.98	41.03	41.09	41.15-	41.21	41.27	41.32	41.38	41.44	41.50
44	41.55+	41.61	41.67	41.73	41.78	41.84	41.90	41.96	42.02	42.07

Phụ lục XIII: BẢNG ĐỔI SANG ARC SIN $\sqrt{\text{Phần trăm}}$ (tiếp theo)

%	0	1	2	3	4	5	6	7	8	9
45	42.13	42.19	42.25-	42.30	42.36	42.42	42.48	42.53	42.59	42.65-
46	42.71	42.76	42.82	42.88	42.94	42.99	43.05-	43.11	43.17	43.22
47	43.28	43.34	43.39	43.45+	43.51	43.57	43.62	43.68	43.74	43.80
48	43.85+	43.91	43.97	44.03	44.08	44.14	44.20	44.25+	44.31	44.37
49	44.43	44.46	44.54	44.60	44.66	44.71	44.77	44.83	44.89	44.94
50	45.00	45.06	45.11	45.17	45.23	45.29	45.34	45.40	45.46	45.52
51	45.57	45.63	45.69	45.75-	45.80	45.86	45.92	45.97	46.03	46.09
52	46.15	46.20	46.26	46.32	46.38	46.43	46.49	46.55	46.61	46.66
53	46.72	46.78	46.83	46.89	46.95+	47.01	47.06	47.12	47.18	47.24
54	47.29	47.35+	47.41	47.47	47.52	47.58	47.64	47.70	47.75+	47.81
55	47.87	47.93	47.98	48.04	48.10	48.16	48.22	48.27	48.33	48.39
56	48.45-	48.50	48.56	48.62	48.68	48.73	48.79	48.85+	48.91	48.97
57	49.02	49.08	49.14	49.20	49.26	49.31	49.37	49.43	49.49	49.54
58	49.60	49.66	49.72	49.78	49.84	49.89	49.95+	50.01	50.07	50.13
59	50.18	50.24	50.30	50.36	50.42	50.48	50.53	50.59	50.65+	50.71
60	50.77	50.83	50.89	50.94	51.00	51.06	51.12	51.18	51.24	51.30
61	51.35+	51.41	51.47	51.53	51.59	51.65-	51.71	51.77	51.83	51.88
62	51.94	52.00	52.06	52.12	52.18	52.24	52.30	52.36	52.42	52.48
63	52.53	52.59	52.65+	52.71	52.77	52.83	52.89	52.95+	53.01	53.07
64	53.13	53.19	53.25-	53.31	53.37	53.43	53.49	53.55-	53.61	53.67
65	53.73	53.79	53.85-	53.91	53.97	54.03	54.09	54.15+	54.21	54.27
66	54.33	54.39	54.45+	54.51	54.57	54.63	54.70	54.76	54.82	54.88
67	54.94	55.00	55.06	55.12	55.18	55.24	55.30	55.37	55.43	55.49
68	55.55+	55.61	55.67	55.73	55.80	55.86	55.92	55.98	56.04	56.11
69	56.17	56.23	56.29	56.35+	56.42	56.48	56.54	56.60	56.66	56.73
70	56.79	56.85+	56.91	56.98	57.04	57.10	57.17	57.23	57.29	57.35+
71	57.42	57.48	57.54	57.61	57.67	57.73	57.80	57.86	57.92	57.99
72	58.05+	58.12	58.18	58.24	58.31	58.37	58.44	58.50	58.56	58.63
73	58.69	58.76	58.82	58.89	58.95+	59.02	59.08	59.15-	59.21	59.28
74	59.34	59.41	59.47	59.54	59.60	59.67	59.74	59.80	59.87	59.93
75	60.00	60.07	60.13	60.20	60.27	60.33	60.40	60.47	60.53	60.60
76	60.67	60.73	60.80	60.87	60.94	61.00	61.07	61.14	61.21	61.27
77	61.34	61.41	61.48	61.55-	61.62	61.68	61.75+	61.82	61.89	61.96
78	62.03	62.10	62.17	62.24	62.31	62.37	62.44	62.51	62.58	62.65+
79	62.72	62.80	62.87	62.94	63.01	63.08	63.15-	63.22	63.29	63.36
80	63.44	63.51	63.58	63.65+	63.72	63.79	63.87	63.94	64.01	64.08
81	64.16	64.23	64.30	64.38	64.45+	64.52	64.60	64.67	64.75-	64.82
82	64.90	64.97	65.05-	65.12	65.20	65.27	65.35-	65.42	65.50	65.57
83	65.65-	65.73	65.80	65.88	65.96	66.03	66.11	66.19	66.27	66.34
84	66.42	66.50	66.58	66.66	66.74	66.81	66.89	66.97	67.05+	67.13
85	67.21	67.29	67.37	67.45+	67.54	67.62	67.70	67.78	67.86	67.94
86	68.03	68.11	68.19	68.28	68.36	68.44	68.53	68.61	68.70	68.78
87	68.87	68.95+	69.04	69.12	69.21	69.30	69.38	69.47	69.56	69.64
88	69.73	69.82	69.91	70.00	70.09	70.18	70.27	70.36	70.45-	70.54
89	70.63	70.72	70.81	70.91	71.00	71.09	71.19	71.23	71.37	71.47
90	71.56	71.66	71.76	71.85+	71.95+	72.05-	72.15-	72.24	72.34	72.44
91	72.54	72.64	72.74	72.84	72.95-	73.05-	73.15+	73.26	73.36	73.46
92	73.57	73.68	73.78	73.89	74.00	74.11	74.21	74.32	74.44	74.55-
93	74.66	74.77	74.88	75.00	75.11	75.23	75.35-	75.46	75.58	75.70
94	75.82	75.94	76.06	76.19	76.31	76.44	76.56	76.69	76.82	76.95-
95	77.08	77.21	77.34	77.48	77.61	77.75+	77.89	78.03	78.17	78.32
96	78.46	78.61	78.76	78.91	79.06	79.22	79.37	79.53	79.69	79.86
97	80.02	80.19	80.37	80.54	80.72	80.90	81.09	81.28	81.47	81.67
98	81.87	82.08	82.29	82.51	82.73	82.96	83.20	83.45+	83.71	83.98

Phụ lục XIII: BẢNG ĐỔI SANG ARC SIN $\sqrt{\text{Phần trăm}}$ (tiếp theo)

%	0	1	2	3	4	5	6	7	8	9
99.0	84.26	84.29	84.32	84.35-	84.38	84.41	84.44	84.47	84.50	84.53
99.1	84.56	84.59	84.62	84.65-	84.68	84.71	84.74	84.77	84.80	84.84
99.2	84.87	84.90	84.93	84.97	85.00	85.03	85.07	85.10	85.13	85.17
99.3	85.20	85.24	85.27	85.31	85.34	85.38	85.41	85.45-	85.48	85.52
99.4	85.56	85.60	85.63	85.67	85.71	85.75-	85.79	85.83	85.87	85.91
99.5	85.95-	85.99	86.03	86.07	86.11	86.15-	86.20	86.24	86.28	86.33
99.6	86.37	86.42	86.47	86.51	86.56	86.61	86.66	86.71	86.76	86.81
99.7	86.86	86.91	86.97	87.02	87.08	87.13	87.19	87.25+	87.31	87.37
99.8	87.44	87.50	87.57	87.64	87.71	87.78	87.86	87.93	88.01	88.10
99.9	88.19	88.28	88.36	88.48	88.60	88.72	88.85+	89.01	89.19	89.43
100.0	90.00									