

THIẾT KẾ BỔ SUNG TRONG CHỌN GIỐNG**Nguyễn Đình Hiền*****Khoa Công nghệ thông tin, Trường Đại học Nông nghiệp Hà Nội****Email: dinhchien535@gmail.com*

Ngày gửi bài: 01.01.2014

Ngày chấp nhận: 27.03.2014

TÓM TẮT

Trong giai đoạn đầu của việc chọn giống, khi có rất nhiều dòng mới cần sàng lọc để giữ lại một số dòng có triển vọng, các nhà nghiên cứu thường dùng thiết kế thí nghiệm bổ sung (hay tăng cường). Đây là kiểu thiết kế đơn giản, dễ bố trí, dễ phân tích nhưng độ chính xác không cao.

Trong bài này chúng tôi giới thiệu sơ qua thiết kế bổ sung và cách phân tích số liệu, sau đó đưa ra các kết quả khi sử dụng các bộ chương trình thống kê Irristat, Cropstat, Minitab, SPSS, Statistica, Jmp và nêu ra nhận xét cá nhân về các kết quả này.

Augmented Design in Plant Breeding**ABSTRACT**

Augmented design is very useful in the early stage of the selection process. In this paper we describe the design, method to analyse data and compare the results given by softwares Excel, Irristat, Cropstat, Minitab, Statistica, SPSS and JMP.

1. GIỚI THIỆU THIẾT KẾ BỔ SUNG

Giả sử có v dòng mới cần thử nghiệm. Đất ở khu thí nghiệm có độ đồng đều cao, được chia thành r khối. Sắp xếp ngẫu nhiên v dòng vào r khối (để thuận tiện thường chọn v là bội số của r), như vậy mỗi khối có v/r dòng thử nghiệm, trong mỗi khối chọn ngẫu nhiên c ô để bố trí c dòng kiểm tra (là các dòng đã biết rõ tính chất và được dùng để so sánh với các dòng thử nghiệm). Có thể vào Irristat utilities, Randomization and layout, Plant breeding design để chọn một Augmented design phù hợp với yêu cầu.

Thiết kế bổ sung tiết kiệm hạt giống vì các dòng thử nghiệm không lặp lại, ô thí nghiệm có

thể chọn loại ô nhỏ, dễ phân tích nhưng độ chính xác của thiết kế này không cao.

Khi so sánh theo phương pháp phân tích phương sai (ANOVA) bậc tự do của sai số $dfE = (r-1)(c-1)$ ít nhất phải là 10 nên số khối:

$$r \geq \frac{10}{(c-1)} + 1$$

Thí dụ: $v = 30$ dòng thử nghiệm, $r = 6$ khối, $c = 3$ dòng kiểm tra thì:

$$\text{Số khối} = 6 (= 10/2 + 1)$$

Vào Irristat để chọn thiết kế bổ sung.

Thí dụ trong [1] 3 dòng kiểm tra, 30 dòng thử nghiệm.

Thiết kế bổ sung trong chọn giống

The screenshot shows the 'Plant Breeding Designs' software window. The 'Design Specification' tab is active. It includes fields for 'Command File' (augmen1.DGN), 'Layout File' (AUGMEN1), and 'Title' (AUGMEN1). The 'Design Type' is set to 'Augmented RCB (max 1000 plots with imbedded RCB less than 800 plots)'. The 'Number of Unreplicated Entries' is 30, 'Number of Replicated Entries' is 3, 'Number of Blocks' is 6, and 'Number of Plots in the Design' is 48. The 'Number of Rows in the Field Layout' is 6. The 'Fieldbook Order' is 'Plot Order', 'Number of Trials' is 1, and 'Random Seed' is 799915.

Kết quả có thể bố trí theo bảng sau, trong đó St, Ci, Wa là 3 dòng kiểm tra (Check) còn 30 dòng thử nghiệm V1, V2, . . . , V30 được chia ngẫu nhiên vào 6 khối.

Bảng 1

B1	B2	B3	B4	B5	B6
St	St	St	St	St	St
V14	Ci	V18	V9	V2	V29
V26	V4	V27	V6	V21	V7
Ci	V15	Vi	Vi	Wa	Vi
V17	V30	V25	Wa	Vi	V1
Wa	V3	V28	C20	V10	Wa
V22	Wa	V5	V11	V8	V12
V13	V24	Wa	V23	V16	V19

Số liệu thu được như sau:

Khoi	Dòng	Nsuat	Khoi	Dòng	Nsuat	Khoi	Dòng	Nsuat
1	St	2972	3	St	2260	5	St	1315
1	14	2405	3	18	2603	5	2	1055
1	26	2865	3	27	2857	5	21	1688
1	Ci	2592	3	Ci	2918	5	Wa	1625
1	17	2572	3	25	2825	5	Ci	1398
1	Wa	2608	3	28	1903	5	10	1293
1	22	2705	3	5	2065	5	8	1253
1	13	2391	3	Wa	3107	5	16	1495
2	St	3122	4	St	3348	6	St	3538
2	Ci	3023	4	9	2268	6	29	2915
2	4	3018	4	6	2148	6	7	3265
2	15	2477	4	Ci	2940	6	Ci	3483
2	30	2955	4	Wa	2850	6	1	3013
2	3	3055	4	20	2670	6	Wa	3400
2	Wa	2477	4	11	3380	6	12	2385
2	24	2783	4	23	2770	6	19	3643

2. CÁCH PHÂN TÍCH

Thiết kế bổ sung được phân tích theo 3 bước:

Bước 1. c dòng kiểm tra (Check varieties) được bố trí trong r khối theo thiết kế khối ngẫu nhiên đủ (RCBD). Gọi kết quả hàng i (check) cột j (block) là X_{ij} . Bảng phân tích phương sai RCBD trong Excel cho ta:

$\bar{x}_i$ là trung bình hàng (Check varieties), $\bar{x}_j$ là trung bình khối, $\bar{\bar{x}}$ là trung bình toàn bộ, số hiệu chỉnh của khối j là $a_j = \bar{x}_j - \bar{\bar{x}}$

Bước 2. Mỗi dòng thử nghiệm (Federer [1] người đưa ra kiểu thí nghiệm bổ sung gọi dòng thử nghiệm là kiểu gen – Genotype) chỉ có mặt một lần trong một khối nên kết quả Y_{ij} (dòng i trong khối j) được hiệu chỉnh bằng cách trừ đi số hiệu chỉnh khối a_j

$$Y_{hc_{ij}} = Y_{ij} - a_j$$

Sai số khi so sánh:

Có nhiều so sánh với nhiều sai số khác nhau:

Hiệu giữa 2 dòng kiểm tra (Check varieties) có sai số:

$$s_c = \sqrt{2MSE/r}$$

Hiệu giữa 2 dòng thử nghiệm (đã hiệu chỉnh) cùng khối có sai số:

$$s_d = \sqrt{2MSE}$$

Hiệu giữa 2 dòng thử nghiệm (đã hiệu chỉnh) khác khối có sai số:

$$s_v = \sqrt{2(c+1)MSE/c}$$

Hiệu giữa dòng thử nghiệm (đã hiệu chỉnh) và một dòng kiểm tra có sai số:

$$s_{vc} = \sqrt{((r+1)(c+1)MSE/rc)}$$

Bước 3. Kết quả

Bảng các số liệu của các dòng kiểm tra viết theo RCBD trong Excel:

Nguồn	Bậc tự do	Tổng bình phương	
Khối	dfR = $r - 1$	$SSR = c \sum_j (\bar{x}_j - \bar{\bar{x}})^2$	
Dòng kiểm tra	dfC = $c - 1$	$SSC = r \sum_i (\bar{x}_i - \bar{\bar{x}})^2$	
Sai số	dfE = $(r-1)(c-1)$	$SSE = SStot - SSR - SSC$	$MSE = SSE/dfE$
Toàn bộ	dfTot = $rc - 1$	$SStot = \sum_i \sum_j (x_{ij} - \bar{\bar{x}})^2$	

	B1	B2	B3	B4	B5	B6	tbinh
Ci	2592	3023	2918	2940	1398	3483	2725.66667
St	2972	3122	2260	3348	1315	3538	2759.16667
Wa	2608	2477	3107	2850	1625	3400	2677.83333
Tbinh xtbj	2724	2874	2761.67	3046	1446	3473.67	2720.88889
Hieu chỉnh	3.11111	153.111	40.7778	325.111	-1274.9	752.778	tbtho

Bảng phân tích phương sai

ANOVA					
Source of Variati	SS	df	MS	F	P-value
Rows	20050.8	2	10025.4	0.11004	0.89686
Columns	6968486	5	1393697	15.2981	0.00021
Error	911027	10	91102.7	MSE	
			se	301.832	
Total	7899564	17			

Bảng các sai số

Diff between 2 check means	$sc = \sqrt{2 * MSE / r}$		174.26
Diff between 2 adj selection mean in the same block	$sd = \sqrt{2 * MSE}$		426.86
Diff between 2 adj selection mean in different block	$sv = \sqrt{2 * (c+1) / c * MSE}$		492.89
Diff between adj selection mean and a check mean	$svc = \sqrt{((c+1) * (r+1) / (r * c)) * MSE}$		376.45

Kết quả đã hiệu chỉnh:

Block	Treat	Yield	sdc	Yhc	Block	Treat	Yield	sdc	Yhc
1	14	2405	3.11	2401.89	4	9	2268	325.1111	1942.889
1	26	2865		2861.89	4	6	2148		1822.889
1	17	2572		2568.89	4	20	2670		2344.889
1	22	2705		2701.89	4	11	3380		3054.889
1	13	2391		2387.89	4	23	2770		2444.889
2	4	3018	153.11	2864.89	5	2	1055	-1274.89	2329.889
2	15	2477		2323.89	5	21	1688		1688.00
2	30	2955		2801.89	5	10	1293		1293.00
2	3	3055		2901.89	5	8	1253		1253.00
2	24	2783		2629.89	5	16	1495		1495.00
3	18	2603	40.78	2562.22	6	29	2915	752.7778	2162.222
3	27	2857		2816.22	6	7	3265		3265.00
3	25	2825		2784.22	6	1	3013		3013.00
3	29	1903		1862.22	6	12	2385		2385.00
3	5	2065		2024.22	6	19	3643		3643.00

Trung bình của các dòng kiểm tra:

St	2759.167
Ci	2725.667
Wa	2677.83

quả hơi khác với **bước 3** và cũng không có bảng các hiệu số kèm sai số như **bước 3**.

- Trong Cropstat 7 phần Tutorial có hướng dẫn cách phân tích Augmented design theo Mixed model analysis (chọn Linear mixed model) kết quả có lúc hoàn toàn giống **bước 3** nhưng có lúc hơi khác với **bước 3** và

cũng không có bảng các hiệu số kèm sai số như **bước 3**.

Các nhận xét trên được rút ra qua việc phân tích thí dụ trong [1], hai thí dụ trong [4] và một số thí dụ khác trong phần Tutorial của Irristat 5 và Cropstat 7.

Sau đây là thí dụ trong [4].

Kết quả trong SAS:

<i>Source</i>	<i>S.E. of Difference</i>	<i>CD(5%)</i>
Control Treatment Means	13.8142	26.8436
Test Treatment in the same Block	23.927	46.4944 $t(0.05,6)$
Test Treatment not in the same Block	26.7512	51.9823 2.446912
Test Treatment and a Control Treatment	20.7214	40.2653

Làm trong Excel				Kết quả theo SAS		
Treat	Y2	adjust	yadj	trt	<i>Adjusted Group Mean Letter</i>	
1	250		256	<i>Control</i>	1 256	DEFG
2	240		228		2 228	G
3	268		247.667		3 247.666667	EFG
4	287		264		4 264	DEF
5	281	+12.917	293.9167	Block 2 test 5	293.916667	CDE
6	300	-0.5833	299.4167	Block3 test 6	299.416667	CD
7	226	-12.333	213.6667	Block1 test 7	213.666667	FG
8	347	-0.5833	346.4167	Block 3 test 8	346.416667	BC
9	311	+12.917	323.9167	Block 2 test 9	323.916667	BC
10	289	-0.5833	288.4167	Block 3 test 10	288.416667	CDE
11	395	-12.333	382.6667	Block 1 test 11	382.666667	AB
12	450	-12.333	437.6667	Block1 test 12	437.666667	A

Theo Minitab 16				Theo JMP 8			Theo SPSS 16		
Least Squares Means for Y2				Level	Sq Mean	Std Error	treat	Mean	Std. Error
Treat	Mean	SE Mean							
1	256.0	9.768		1	256.00000	9.768146	1	256.000	9.768
2	228.0	9.768		2	228.00000	9.768146	2	228.000	9.768
3	247.7	9.768		3	247.66667	9.768146	3	247.667	9.768
4	264.0	9.768		4	264.00000	9.768146	4	264.000	9.768
5	293.9	18.275		5	293.91667	18.274527	5	293.917	18.275
6	299.4	18.275		6	299.41667	18.274527	6	299.417	18.275
7	213.7	18.275		7	213.66667	18.274527	7	213.667	18.275
8	346.4	18.275		8	346.41667	18.274527	8	346.417	18.275
9	323.9	18.275		9	323.91667	18.274527	9	323.917	18.275
10	288.4	18.275		10	288.41667	18.274527	10	288.417	18.275
11	382.7	18.275		11	382.66667	18.274527	11	382.667	18.275
12	437.7	18.275		12	437.66667	18.274527	12	437.667	18.275

Augmented design			Mixed model analysis			General linear model		
theo Irristat 5 và Cropstat7			TheoCropstat 7			Theo Statistica 8		
ENTRY	Y2		LSMean	Std. Error		Treat	Y2	Y2
ENTRY ADJUSTED						Mean	Std.Err.	
1	256.0		1	256.0	11.16	1	256.0000	9.76815
2	228.0		2	228.0	11.16	2	228.0000	9.76815
3	247.7		3	247.7	11.16	3	247.6667	9.76815
4	264.0		4	264.0	11.16	4	264.0000	9.76815
5	293.9		5	288.1	18.48	5	293.9167	18.27453
6	299.4		6	299.7	18.48	6	299.4167	18.27453
7	213.7		7	219.2	18.48	7	213.6667	18.27453
8	346.4		8	346.7	18.48	8	346.4167	18.27453
9	323.9		9	318.1	18.48	9	323.9167	18.27453
10	288.4		10	288.7	18.48	10	288.4167	18.27453
11	382.7		11	388.2	18.48	11	382.6667	18.27453
12	437.7		12	443.2	18.48	12	437.6667	18.27453
MEANS	298.5							

TÀI LIỆU THAM KHẢO

- [1] Roger G Petersen (1994). Agricultural field experiment. Marcel Dekker Inc.
- [2] Các phần mềm thống kê Irristat 5, Cropstat 7 của Viện lúa quốc tế IRRI có thể tải miễn phí trên Internet
- [3] Các phần mềm MiniTab 16, Statistica 8, SPSS 16, JMP 8 là các bản Demo tải miễn phí trên Internet.
- [4] Analysis of augmented designs using SAS. Tài liệu của Viện nghiên cứu thống kê trong nông nghiệp Ấn độ (IASRI).