THIẾT KẾ BỔ SUNG TRONG CHỌN GIỐNG

Nguyễn Đình Hiền

Khoa Công nghệ thông tin, Trường Đại học Nông nghiệp Hà Nội

Email: dinhhien535@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 descibe 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 \mathbf{v} dòng mới cần thử nghiệm. Đất ở khu thí nghiệm có độ đồng đều cao, được chia thành \mathbf{r} khối. Sắp xếp ngẫu nhiên \mathbf{v} dòng vào \mathbf{r} khối (để thuận tiện thường chọn \mathbf{v} là bội số của \mathbf{r}), như vậy mỗi khối có $\mathbf{v/r}$ dòng thử nghiệm, trong mỗi khối chọn ngẫu nhiên \mathbf{c} ô để bố trí \mathbf{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:

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

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

So' kho' = 6 (= 10/2 + 1)

Vào Irristat để chọn thiết kế bổ sung.

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

🚟 Plant Breeding Designs	B			
Design Specification				
👝 Open				ок
Command File: augmen1.DG	N	Layout File:	AUGMEN1	
Title: AUGMEN1				
Design Type: Augmented RCE) (max 1000 plots with im	bedded RCB les	ss than 800 plots)	•
Number of Unreplicated Entries:	30 🛨 Nu	mber of Replicat	ted Entries: 3	•
Number of Blocks:	6			
Number of Plots in the Design:	48 Nu	mber of Rows in	n the Field Layout: 6	•
Fieldbook Order: Plot Order	Number of Trials:	Rand	om Seed: 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.

		Dui	-8 -		
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

Bång 1

Số liệu thu được như sau:

Khoi	Dong	Nsuat		Khoi	Dong	Nsuat	Khoi	Dong	Nsuat
1	St	2972	Ĺ	3	St	2260	5	St	1315
1	14	2405	ĺ	З	18	2603	5	2	1055
1	26	2865		З	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		З	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:

 $\overline{x_i}$ là trung bình hàng (Check varieties), $\overline{x_j}$ là trung bình khối, $\overline{\overline{x}}$ là trung bình toàn bộ, số hiệu chỉnh của khối j là $\mathbf{a_j} = \overline{x_j} - \overline{\overline{x}}$

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 \mathbf{Y}_{ij} (dòng i trong khối j) được hiệu chỉnh bằng cách trừ đi số hiệu chỉnh khối \mathbf{a}_i

 $Yhc_{ij} = Y_{ij} - a_j$ 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 giưa dòng thử nghiêm (đã hiệu chỉnh) và một dòng kiểm tra có sai số:

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

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	$\text{SSR} = c \sum_{j} (\overline{x}_{j} - \overline{\overline{x}})^{2}$	
Dòng kiểm tra	dfC = c- 1	SSC = $r \sum_{i} (\overline{x}_{i} - \overline{\overline{x}})^{2}$	
Sai số	dfE = (r-1)(c-1)	SSE = SStot – SSR - SSC	MSE = SSE/dfE
Toàn bộ	dfTot = rc - 1	SSTot = $\sum_{i} \sum_{j} (x_{ij} - \overline{\overline{x}})^2$	

	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 chinh	3.11111	153.111	40.7778	325.111	-1274.9	752.778	tbtbo

	ANOVA					
So	urce 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 phân tích phương sai

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)						
Diff between 2 adj selection mean in differentblock sv= sqrt(2*(c+1)/c*MSE)						
Diff between adj selection mea	n and a check mean svc = sqrt((c+1)*(r+1)/(r*c)*N	VISE) 376.45				

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

Kết quả đã hiệu chỉnh:

Trung bình của các dòng kiểm tra:

St	2759.167
Ci	2725.667
Wa	2677.83

3. KẾT QUẢ PHÂN TÍCH THEO MỘT SỐ PHẦN MỀM THỐNG KÊ THÔNG DỤNG

Bước 3 cho kết quả phân tích theo cách làm truyền thống do Federer [1] đề xuất và được thực hiện từng bước bằng phần mềm Excel.

Dùng phần mềm Irristat 5 có thể tạo được thiết kế bổ sung (Irristat utilities, Randomization and layout, Plant breeding design), vào số liệu sau đó chọn Analysis, Single site analysis, kiểu thiết kế Augmented design, được kết quả phân tích, bảng phân tích phương sai, sai số và các giá trị Y đã hiệu chỉnh như **bước 3.** - Dùng phần mềm Cropstat 7 có thể tạo được thiết kế bổ sung như Irristat 5 còn trong Single site analysis chọn thiết kế Regular controls in RCBD được kết quả như **bước 3.**

- Dùng MiniTab 16, Statistica 8, JMP 8 chọn Anova General linear model với mô hình: Y = Block + Treatment.

Tìm ES means, được các Y đã hiệu chỉnh, bảng phân tích phương sai tương tự như **bước 3** nhưng không có bảng các hiệu số kèm sai số như **bước 3**.

- Dùng SPSS 16 Anova General linear model được ES means một số trường hợp cho kết

Block	Treat	¥2	Block	Treat	¥2			
1	1	250	3	1	258			
1	2	240	3	2	224			
1	3	268	3	3	238			
1	4	287	З	4	278			
1	7	226	З	8	347			
1	11	395	З	6	300			
1	12	450	з	10	289			
2	1	260						
2	2	220	1,2,3,4 là 4 dòng kiểm tra					
2	3	237	5,6,7,8	,9,10,11	,12 là 8	dòng thử nghiệm		
2	4	227						
2	5	281						
2	9	311						
3	1	258						
З	2	224						
з	3	238						
з	4	278						
3	8	347						
3	6	300						
3	10	289						

Kết quả trong Excel:

	B1	B2	B3	Tbinh						
1	250	260	258	256.00			ANOVA			
2	240	220	224	228.00		Sou	rce of Varia	SS	df	MS
3	268	237	238	247.67		Kiểm tra	Rows	2150.25	3	716.75
4	287	227	278	264.00		Block	Columns	1277.167	2	638.5833
							Error	1717.5	6	286.25
xtb	261.25	236	249.5	248.9167			Total	5144.917	11	
adj	12.33333	-12.9167	0.583333							
		LSD khi so	sanh 2 tru	ng binh						
		2 Trung bi	nh cua 2 dà	òng kiểm tr	a (Ch	neck hay Co	ontrol)		33.80224	
		2 trung bii	nh cua 2 dò	ng thử ngi	niệm	cùng Bloc	k		58.54719	
		2 trung bii	nh 2 dòng t	hử nghiện	n khá	c block			65.45775	
		trung binł	n cua thử n	ghiệm so v	oi tru	ung bình củ	la dòng kiể	m tra	53.44603	

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:

Source	S.E. of Difference	CD(5%)	
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	¥2	adjust	yadj	trt	Adjuste trt Mea			l Group Letter	
1	250		256			1	256	DEFG	
2	240		228			2	228	G	
3	268		247.667	Control		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	А	

Theo Minitab 16		Theo JMP	Theo SPSS 16			
Least Squares Means for Y2	Level	Sq Mean	Std Error	treat	Mean	Std. Error
Treat Mean SE Mean	1	256.00000	9.768146	1	256.000	9.768
1 256.0 9.768	2	228.00000	9.768146	2	228.000	9,768
2 228.0 9.768	3	247.66667	9.768146	3	247.667	9,768
3 247.7 9.768	4	264 00000	9,768146	4	264 000	9 768
4 264.0 9.768	5	293 91667	18 274527	5	293 917	18 275
5 293.9 18.275	6	200.01001	18 274527	96	200.011	18 275
6 299.4 18.275	7	212 66667	10.274527	7	212 667	19 275
7 213.7 18.275		213.00007	10.274527	0	213.007	10.275
8 346.4 18.275	8	340.41007	18.274527	°	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 27 45 27	12	437.667	18.275

Augmented design theo Irristat 5 va Cropstat7			Mixed model analysis TheoCropstat 7			General linear model Theo Stattstica 8			
7	256.0	-	2	228.0	11.16	1	256.0000	9.76815	
3	247.7	-	3	247.7	11.16	2	228.0000	9.76815	
4	264.0	-	4	264.0	11.16	3	247.6667	9.76815	
5	293.9		5	288.1	18.48	4	264.0000	9.76815	
6	299.4		6	299.7	18.48	- 5	293.9167	18.27453	
7	213.7		2	219.2	18.48	6	299.4167	18.27453	
8	346.4			346 7	18 48	7	213.6667	18.27453	
10	323.9 288 4		å	319 1	19 49	8	346.4167	18.27453	
			10	200 7	10.40	9	323.9167	18.27453	
11	382.7		10	200.7	10.40	10	288.4167	18.27453	
12	437.7	-		388.2	18.48	11	382,6667	18.27453	
MEANS	298.5		12	443.2	18.48	12	437.6667	18.27453	

TÀI LIỆU THAM KHẢO

- [1] Roger G Petersen (1994). Agricultural field expriment. 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).