

THIẾT KẾ CROSSOVER TRONG CÁC KIỂM ĐỊNH SINH HỌC

Nguyễn Đình Hiền

Bộ môn Công nghệ phần mềm

Trong bài thiết kế AB/BA trong kiểm định thuốc chúng tôi đã giới thiệu thiết kế AB/BA để kiểm định 2 loại thuốc qua 2 giai đoạn. Trong bài này chúng tôi sẽ giới thiệu thiết kế Crossover.

1 Thiết kế Crossover 3 cách xử lý (Treatment), 3 giai đoạn (period)

Thí dụ cụ thể trong [1]: 3 loại thức ăn A, B, C thử nghiệm trên 12 con cừu theo 6 dãy (sequence) ABC, BCA, CAB, ACB, BAC, CBA mỗi dãy 2 con (units) mỗi thức ăn dùng trong 1 giai đoạn (period) 3 tháng. Biến kết quả là mức tăng trọng

PERIODS	ANIMAL NUMBERS											
	53	54	58	75	81	97	72	79	106	84	89	70
1	A 72	B 75	C 75	A 64	B 80	C 74	A 58	B 64	C 72	B 76	A 61	C 71
2	B 73	C 78	A 77	C 68	A 72	B 76	C 62	A 56	B 69	C 79	B 50	A 72
3	C 77	A 70	B 73	B 71	C 80	A 70	B 67	C 60	A 66	A 65	C 60	B 75

Để phân tích thiết kế Crossover này trước hết phải viết lại số liệu

day	units	periods	treat	residual	yield
1	1	1	1	1	72
1	1	2	2	1	73
1	1	3	3	2	77
2	2	1	2	1	75
2	2	2	3	2	78
2	2	3	1	3	70
3	3	1	3	1	75
3	3	2	1	3	77
3	3	3	2	1	73
4	4	1	1	1	64
4	4	2	3	1	68
4	4	3	2	3	71
5	5	1	2	1	80
5	5	2	1	2	72
5	5	3	3	1	80
6	6	1	3	1	74
6	6	2	2	3	76
6	6	3	1	2	70

day	units	periods	treat	residual	yield
4	7	1	1	1	58
4	7	2	3	1	62
4	7	3	2	3	67
5	8	1	2	1	64
5	8	2	1	2	56
5	8	3	3	1	60
6	9	1	3	1	72
6	9	2	2	3	69
6	9	3	1	2	66
1	10	1	2	1	76
1	10	2	3	2	79
1	10	3	1	3	65
2	11	1	1	1	61
2	11	2	2	1	50
2	11	3	3	2	60
3	12	1	3	1	71
3	12	2	1	3	72
3	12	3	2	1	75

Trong đó residual là ảnh hưởng tồn dư (Carry over) với cách viết: ở giai đoạn 1 ghi 1, ở các giai đoạn sau căn cứ vào treat ở giai đoạn trước: A ghi 1, B ghi 2, C ghi 3

1.1 IASRI (Indian Agricultural Research Institut) đưa ra mô hình

$$\text{Yield} = \mu + \text{units} + \text{periods} + \text{treat} + \text{residual} + \text{sai số ngẫu nhiên} \quad (1)$$

Mô hình này có thể phân tích trong SPSS, MiniTab, Jmp, SAS

Sau đây là Syntax trong SPSS

```
UNIANOVA
  yield BY units periods treat residual
  /METHOD = SSTYPE(3)
  /INTERCEPT = INCLUDE
  /EMMEANS = TABLES(treat) COMPARE ADJ(LSD)
  /EMMEANS = TABLES(residual) COMPARE ADJ(LSD)
  /CRITERIA = ALPHA(.05)
  /DESIGN = units periods treat residual .
```

Một số kết quả

Tests of Between-Subjects Effects

Dependent Variable: yield

Source		Type III Sum of Squares	df	Mean Square	F	Sig.
Intercept	Hypothesis	119427.042	1	119427.042	1366.237	.000
	Error	1050.600	12.019	87.413 ^a		
period	Hypothesis	45.729	2	22.865	1.835	.188
	Error	224.333	18	12.463 ^b		
treat	Hypothesis	191.633	2	95.817	7.688	.004
	Error	224.333	18	12.463 ^b		
units	Hypothesis	1328.800	11	120.800	9.693	.000
	Error	224.333	18	12.463 ^b		
resid	Hypothesis	55.500	2	27.750	2.227	.137
	Error	224.333	18	12.463 ^b		

a. .692 MS(units) + .308 MS(Error)

b. MS(Error)

Dependent Variable: yield

treat	Mean	Std. Error	95% Confidence Interval	
			Lower Bound	Upper Bound
1	66.917	1.019	64.776	69.058
2	72.000	1.248	69.378	74.622
3	72.708	1.248	70.086	75.331

Estimates

Dependent Variable: yield

resid	Mean	Std. Error	95% Confidence Interval	
			Lower Bound	Upper Bound
1	67.917	1.019	65.776	70.058
2	71.667	1.611	68.281	75.052
3	72.042	1.611	68.656	75.427

Chương trình SAS

```
proc glm data=cod;
class units periods treat residual;
model yield = units periods treat residual/ss3;
lsmeans treat/pdiff adjust =tukey;
lsmeans residual/pdiff adjust =tukey;
run;
Một số kết quả
```

The GLM Procedure

Dependent Variable: yield

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	17	1625.666667	95.627451	7.67	<.0001
Error	18	224.333333	12.462963		
Corrected Total	35	1850.000000			

R-Square	Coeff Var	Root MSE	yield Mean
0.878739	5.067405	3.530292	69.66667

Source	DF	Type III SS	Mean Square	F Value	Pr > F
units	11	1328.800000	120.800000	9.69	<.0001
periods	2	45.729167	22.864583	1.83	0.1883
treat	2	191.633333	95.816667	7.69	0.0039
residual	2	55.500000	27.750000	2.23	0.1368

Least Squares Means
Adjustment for Multiple Comparisons: Tukey-Kramer

treat	yield LSMEAN	Number
1	66.9166667	1
2	72.0000000	2
3	72.7083333	3

Least Squares Means for effect treat
Pr > |t| for H0: LSMean(i)=LSMean(j)
Dependent Variable: yield

i/j	1	2	3
1		0.0144	0.0056
2	0.0144		0.8995
3	0.0056	0.8995	

1.2 Một số tác giả đưa thêm biến dãy (seq) vào mô hình

$\text{Yield} = \mu + \text{seq} + \text{units}(\text{seq}) + \text{periods} + \text{treat} + \text{residual} + \text{sai số ngẫu nhiên}$ (2)

Trong SPSS

```
UNIANOVA yield BY day periods treat residual units
/RANDOM=units
```

```

/METHOD=SSTYPE(3)
/INTERCEPT=INCLUDE
/EMMEANS=TABLES(periods)
/EMMEANS=TABLES(treat)
/CRITERIA=ALPHA(.05)
/DESIGN=day units(day) periods treat residual.

```

Chương trình SAS

```

proc glm data=cod;
class day units periods treat residual;
model yield = day units(day) periods treat residual/ss3;
TEST H = day E =units(day) /HTYPE=1 ETYPE=1;
lsmeans treat/pdiff adjust =tukey;
lsmeans residual/pdiff adjust =tukey;
run;

```

Có thể dùng chương trình

```

proc MIXED data=cod;
class day units periods treat residual;
model yield = day periods treat residual;
RANDOM units(day);
lsmeans day treat residual;
run;

```

hoặc

```

proc MIXED data=cod;
class day units periods treat residual;
model yield = day periods treat residual;
REPEATED / SUBJECT=units TYPE=CS R;
lsmeans day treat residual;
run;

```

1.3 Một số tác giả bỏ qua residual và thêm tương tác period* treat

Yield = μ + seq + units(seq) + periods + treat + period *treat + sai số ngẫu nhiên (3)

(Vì quan niệm tương tác period*treat thể hiện ảnh hưởng tồn dư residual)

Thí dụ tệp số liệu có 3 sequence, 4 subject, 3 period, 3 treatment

SEQ	SUBJECT	PERIOD	TREAT	DURATION	SEQ	SUBJECT	PERIOD	TREAT	DURATION
1	1	1	1	1.5	1	2	2	2	2.6
1	1	2	2	2.2	1	2	3	3	3.1
1	1	3	3	3.4	1	3	1	1	1.6
1	2	1	1	2	1	3	2	2	2.7

1	3	3	3	3.2
1	4	1	1	1.1
1	4	2	2	2.3
1	4	3	3	2.9
2	1	1	2	2.5
2	1	2	3	3.5
2	1	3	1	1.9
2	2	1	2	2.8
2	2	2	3	3.1
2	2	3	1	1.5
2	3	1	2	2.7
2	3	2	3	2.9
2	3	3	1	2.4
2	4	1	2	2.4
2	4	2	3	2.6

2	4	3	1	2.3
3	1	1	3	3.3
3	1	2	1	1.9
3	1	3	2	2.7
3	2	1	3	3.1
3	2	2	1	1.6
3	2	3	2	2.5
3	3	1	3	3.6
3	3	2	1	2.3
3	3	3	2	2.2
3	4	1	3	3
3	4	2	1	2.5
3	4	3	2	2

Trong SPSS

UNIANOVA

dur BY seq period treat sub

/RANDOM = sub

/METHOD = SSTYPE(1)

/INTERCEPT = INCLUDE

/EMMEANS=TABLES(treat)

Một số kết quả

/EMMEANS=TABLES(period)

/ Test = seq vs sub (seq)

/CRITERIA = ALPHA(.05)

/DESIGN = seq period treat period*treat
sub(seq)

1. treat

Dependent Variable: dur				
treat	Mean	Std. Error	95% Confidence Interval	
			Lower Bound	Upper Bound
1	1.883	.098	1.678	2.089
2	2.467	.098	2.261	2.672
3	3.142	.098	2.936	3.347

Test Results

Dependent Variable: dur					
Source	Sum of Squares	df	Mean Square	F	Sig.
Contrast	.234	2	.117	1.573	.260
Error ^a	.669	9	.074		

a. sub(seq)

Tests of Between-Subjects Effects

Dependent Variable: dur

Source	Type I Sum of Squares	df	Mean Square	F	Sig.
Intercept	224.500	1	224.500	3019.431	.000
seq	.669	9	.074 ^a	1.573	.260
period	.234	2	.117	.075	.928
treat	.669	9	.074 ^a	41.513	.000
period *	.017	2	.009	2.830	.085
treat	2.063	18	.115 ^b		
sub(seq)	9.517	2	4.759		
	2.063	18	.115 ^b		
	.649	2	.324		
	2.063	18	.115 ^b		
	.669	9	.074		
	2.063	18	.115 ^b		

a. MS(sub(seq))

b. MS(Error)

Trong SAS có thể dùng Procedure GLM hay Mixed

```
PROC GLM;
  CLASS SEQUENCE SUBJECT PERIOD TREATMENT;
  MODEL DURATION = SEQUENCE SUBJECT(SEQUENCE) TREATMENT PERIOD
    TREATMENT*PERIOD;
  TEST H = SEQUENCE E = SUB(SEQUENCE) / HTYPE=1 ETYPE=1;
  LSMEANS TREAT / PDIFF CL E;
RUN;
```

Một số kết quả trong SAS

Dependent Variable: DURATION					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	17	11.08638889	0.65214052	5.69	0.0003
Error	18	2.06333333	0.11462963		
Corrected Total	35	13.14972222			
R-Square					
0.843089					
Coeff Var					
13.55786					
Root MSE					
0.338570					
DURATION Mean					
2.497222					

Source	DF	Type III SS	Mean Square	F Value	Pr > F
SEQUENCE	0	0.00000000	.	.	.
SUBJECT(SEQUENCE)	9	0.66916667	0.07435185	0.65	0.7425
TREATMENT	2	9.51722222	4.75861111	41.51	<.0001
PERIOD	2	0.01722222	0.00861111	0.08	0.9279
PERIOD*TREATMENT	2	0.64888889	0.32444444	2.83	0.0853

Tests of Hypotheses Using the Type I MS for SUBJECT(SEQUENCE) as an Error Term					
Source	DF	Type I SS	Mean Square	F Value	Pr > F
SEQUENCE	2	0.23388889	0.11694444	1.57	0.2595

The GLM Procedure				
Least Squares Means				
Least Squares Means for effect TREATMENT				
Pr > t for H0: LSMean(i)=LSMean(j)				
Dependent Variable: DURATION				
i/j	1	2	3	
1		0.0005	<.0001	
2	0.0005		0.0001	
3	<.0001	0.0001		

TREATMENT	DURATION LSMEAN	95% Confidence Limits	
1	1.883333	1.677996	2.088671
2	2.466667	2.261329	2.672004
3	3.141667	2.936329	3.347004

Least Squares Means for Effect TREATMENT				
i	j	Difference Between Means	95% Confidence Limits for LSMean(i)-LSMean(j)	
1	2	-0.583333	-0.873724	-0.292943
1	3	-1.258333	-1.548724	-0.967943
2	3	-0.675000	-0.965391	-0.384609

2 Thiết kế Crossover nhiều cách xử lý

Khi có 2 cách xử lý, 2 giai đoạn thì thường chỉ dùng 2 dãy thử nghiệm 2 giai đoạn AB / BA, tuy nhiên cũng có trường hợp dùng thiết kế 2 dãy 3 giai đoạn ABA, BAB.

Khi có n (n > 2) cách xử lý thì có nhiều dãy thử nghiệm trong đó mỗi cách xử lý có mặt một số lần ở mỗi dãy và có mặt một số lần ở mỗi giai đoạn, thí dụ 3 cách xử lý A, B, C 3 giai đoạn có thể chọn thiết kế 6 dãy thử nghiệm ABC, BCA, CAB, ACB, BAC, CBA mỗi cách xử lý có mặt một lần trong một dãy và hai lần trong một giai đoạn.

Ảnh hưởng tồn dư và tương tác giữa ảnh hưởng tồn dư và giai đoạn cũng có nhiều vấn đề phức tạp do đó thường chọn các dãy thử nghiệm đơn giản, dễ sử dụng, có thể lập mô hình để xử lý trên máy tính qua đó tìm ra cách đánh giá hợp lý và công bằng các cách xử lý.

Một trong các dãy hay được chọn là các dãy **đồng đều**

Đồng đều trong dãy thử nghiệm (Uniform within sequences): mỗi cách xử lý có mặt một số lần như nhau trong mỗi dãy thử nghiệm

Đồng đều trong giai đoạn (Uniform within periods): mỗi cách xử lý có mặt một số lần như nhau trong mỗi giai đoạn

Nếu vừa đồng đều trong dãy vừa đồng đều trong giai đoạn thì gọi tắt là đồng đều **(uniform)**.

Nếu thiết kế cân bằng về một khía cạnh nào đó thì rất tốt, thí dụ cân bằng đối với các ước lượng của các cách xử lý (phương sai của các ước lượng như nhau, phương sai của hiệu 2 cách xử lý như nhau đối với mọi cặp cách xử lý . . .) cân bằng đối các ước lượng của ảnh hưởng tồn dư . . .

Thông thường hay đòi hỏi **cân bằng** đối với ảnh hưởng tồn dư bậc nhất (**balanced with respect to first-order carryover effects**) theo nghĩa: số lần một cách xử lý đứng trước một cách xử lý khác là như nhau đối với mọi cách xử lý, nếu số lần đó bằng số lần đứng trước bản thân nó thì có **cân bằng mạnh (Strongly Balanced)**.

Thí dụ 3 cách xử lý 3 giai đoạn, thiết kế hay dùng là 3 dãy ABC, BCA, CAB nhưng để cân bằng có thể dùng 6 dãy xử lý xếp trong 2 ô vuông Latinh

	Ô vuông la tinh thứ nhất				Ô vuông la tinh thứ nhất		
	Dãy 1	Dãy 2	Dãy 3		Dãy 4	Dãy 5	Dãy 8
Period 1	A	B	C		A	B	C
Period 2	B	C	A		C	A	B
Period 3	C	A	B		B	C	A

Mỗi cách xử lý A, B, C có mặt 2 lần trong mỗi giai đoạn (Uniform within periods), có mặt 1 lần trong mỗi dãy (Uniform within periods), A đứng trước B 2 lần, đứng trước C 2 lần, B đứng trước A 2 lần, đứng trước C 2 lần, C đứng trước A 2 lần, trước B 2 lần (Balanced) nhưng không strongly balanced

Thiết kế 4 dãy 4 giai đoạn 2 cách xử lý (Viết tắt là ABBA, BAAB, AABB, BBAA) là thiết kế Strongly balanced.

[Design 10]	Period 1	Period 2	Period 3	Period 4
Sequence	A	B	B	A

ABBA				
Sequence BAAB	B	A	A	B
Sequence AABB	A	A	B	B
Sequence BBAA	B	B	A	A

Thiết kế ô vuông latin 4 x 4

LATINH1			
A	B	C	D
B	C	D	A
C	D	A	B
D	A	B	C

LATINH2			
A	B	C	D
B	D	A	C
C	A	D	B
D	C	B	A

Coi hàng là giai đoạn, cột là dãy thì latin1 uniform (mỗi chữ cái A, B, C, D có mặt một lần trong 1 cột và 1 lần trong 1 hàng) nhưng không balanced.

Latinh2 vừa uniform vừa balanced do đó hay được chọn trong thiết kế crossover

Sau đây là một số loại thiết kế khác

<u>Uniform within Sequences</u>	<u>Uniform within Periods</u>	<u>Balanced</u>	<u>Strongly Balanced</u>	<u>Examples</u>
no	no	no	no	AAB ABB, ABCC BCAA
yes	no	no	no	ABB BAB, ABC CBA
no	yes	no	no	ABCC BCAA CABB
no	no	yes	no	ABAA BAAB
no	no	yes	yes	AABBA BAABB
yes	yes	no	no	ABC BCA CAB
yes	no	yes	no	AABA ABAA
no	yes	yes	no	ABA BAB
yes	no	yes	yes	AABBA ABBAA
no	yes	yes	yes	ABB BAA, AB BA AA BB
yes	yes	yes	no	AB BA
yes	yes	yes	yes	ABBA BAAB AABB BBAA

Về mô hình để phân tích thường dùng mô hình tuyến tính

$$y_{ijk} = \mu + \alpha_i + b_{ij} + \gamma_k + \tau_{d(i,k)} + \lambda_{c(i,k-1)} + e_{ijk}$$

μ là trung bình chung, α_i ảnh hưởng của dãy thử nghiệm (sequence) i , b_{ij} ảnh hưởng của cá thể (subject) j trong dãy thử nghiệm i , γ_k ảnh hưởng của giai đoạn (period) k , $\tau_{d(i,k)}$ ảnh hưởng của cách xử lý (treatment) d trong dãy thử nghiệm i ở giai đoạn k , $\lambda_{c(i,k-1)}$ ảnh hưởng dư thừa (carryover) c trong dãy thử nghiệm i ở giai đoạn $k - 1$.

Có thể phân tích mô hình này trong SPSS, SAS.

3 Thiết kế Williams

Thiết kế Williams là thiết kế vừa uniform vừa balanced được xây dựng từ một ô vuông Latinh đặc

biệt [5]. Nếu n chẵn thì chỉ cần một ô vuông Latinh, nếu n lẻ thì cần ghép 2 ô vuông Latinh.

Thí dụ Williams 3 cách xử lý, 3 giai đoạn, 6 dãy gồm 2 ô vuông Latinh 3x3

A	B	C
B	C	A
C	A	B

C	A	B
B	C	A
A	B	C

Thí dụ Williams 4 cách xử lý, 4 giai đoạn, 4 dãy, chọn 1 trong 2 ô vuông 4 x 4

Ô vuông 4 x 4 thứ nhất

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

ô vuông 4 x 4 thứ hai

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

Thí dụ Williams 5 cách xử lý, 5 giai đoạn, 10 dãy gồm 2 ô vuông 5 x 5

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

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

Thí dụ Williams 6 cách xử lý, 6 giai đoạn, 6 dãy gồm 1 ô vuông 6 x 6

A	B	C	D	E	F
C	D	E	F	A	B
B	C	D	E	F	A
E	F	A	B	C	D
F	A	B	C	D	E
D	E	F	A	B	C

4 Một số bài toán liên quan đến thiết kế crossover

Thiết kế Crossover được dùng chủ yếu trong các kiểm định thuốc, các kiểm định sinh học, nông nghiệp. Trong ngành công nghệ thực phẩm thiết kế crossover được dùng để đánh giá chất lượng rượu, chất lượng các loại thực phẩm.

Thí dụ 1 Tám ủy viên trong ban giám khảo đánh giá 4 loại rượu A, B, C, D theo thiết kế vừa uniform vừa balanced sau:

	GK1	GK2	GK3	GK4	GK5	GK6	GK7	GK8
GDOAN1	B	C	D	A	B	A	C	D
GDOAN2	A	B	C	D	A	C	D	B
GDOAN3	C	D	A	B	D	B	A	C
GDOAN4	D	A	B	C	C	D	B	A

Thí dụ 2 36 khách hàng cho nhận xét về 3 loại thức ăn. Chọn ngẫu nhiên 6 khách hàng để nếm 3 loại thức ăn theo 6 dãy

Dãy 1	Dãy 2	Dãy 3	Dãy 4	Dãy 5	Dãy 6
A	B	C	A	C	B
B	C	A	C	B	A
C	A	B	B	A	C

Tài liệu tham khảo:

- 1- Indian Agricultural Research Institut IASRI. Crossover Design Example.
 - 2- Stat 509 Design and analysis of clinical trials. The Pennsylvania State University.
 - 3- Crossover study. Wikipedia the free encyclopedia
 - 4- Williams EJ (1949). Experimental Designs Balanced for the Estimation of Residual effects of Treatments." Australian Journal of Scientic Research, 2(3), 149-168.
 - 5-Jones and Kenward (2003) Design and Analysis of Cross-Over Trials, Second Edition. Chapman and Hall
-