Lecture 13: Blocking and Confounding in 2^k Design

Montgomery: Chapter 7

Randomized Complete Block 2^k Design

- There are n blocks
- Within each block, all treatments (level combinations) are conducted.
- Run order in each block must be randomized
- Analysis follows general block factorial design
- When k is large, cannot afford to conduct all the treatments within each block.
  - Other blocking strategy should be considered.

Filtration Rate Experiment (Revisited)

factor				
A	B	C	D	original response
$-$	$-$	$-$	$-$	45
$+$	$-$	$-$	$-$	71
$-$	$+$	$-$	$-$	48
$+$	$+$	$-$	$-$	65
$-$	$-$	$+$	$-$	68
$+$	$-$	$+$	$-$	60
$-$	$+$	$+$	$-$	80
$+$	$+$	$+$	$-$	65
$-$	$-$	$-$	$+$	43
$+$	$-$	$-$	$+$	100
$-$	$+$	$-$	$+$	45
$+$	$+$	$-$	$+$	104
$-$	$-$	$+$	$+$	75
$+$	$-$	$+$	$+$	86
$-$	$+$	$+$	$+$	70
$+$	$+$	$+$	$+$	96

- Suppose there are two batches of raw material.
  - Each batch can be used for only 8 runs.
  - It is known these two batches are very different.
  - Blocking should be employed to eliminate this variability.
- How to select 8 treatments (level combinations, or runs) for each block?

2^2 Design with Two Blocks

- There are two factors (A, B) each with 2 levels
  - Two blocks (b_1, b_2) each contain two runs (treatments)
  - Since b_1 and b_2 are interchangeable, there are three possible blocking scheme:

			blocking scheme		
A	B	response	1	2	3
$-$	$-$	y_{--}	b_1	b_1	b_2
$+$	$-$	y_{+-}	b_1	b_2	b_1
$-$	$+$	y_{-+}	b_2	b_1	b_1
$+$	$+$	y_{++}	b_2	b_2	b_2

Comparing Blocking Schemes

- Scheme 1:
  - block effect: $b = \bar{y}_{b_2} - \bar{y}_{b_1} = \frac{1}{2}(-y_{--} - y_{+-} + y_{-+} + y_{++})$
  - main effect: $B = \frac{1}{2}(-y_{--} - y_{+-} + y_{-+} + y_{++})$
  - B and b are not distinguishable, or, confounded.
- Scheme 2:
  - block effect: $b = \bar{y}_{b_2} - \bar{y}_{b_1} = (-y_{--} + y_{+-} - y_{-+} + y_{++})/2$
  - main effect: $A = (-y_{--} + y_{+-} - y_{-+} + y_{++})/2$
  - A and b are not distinguishable, or confounded.
- Scheme 3:
  - block effect: $b = \bar{y}_{b_2} - \bar{y}_{b_1} = (y_{--} - y_{+-} - y_{-+} + y_{++})/2$
  - interaction: $AB = (y_{--} - y_{+-} - y_{-+} + y_{++})/2$
  - AB and b become indistinguishable, or confounded.

2^k Design with Two Blocks via Confounding

- The reason for confounding: the block arrangement matches the contrast of some factorial effect.
- Confounding makes the effect **Inestimable**.
- **Question: which scheme is the best (or causes the least damage)?**
- **Confound blocks with the effect (contrast) of the highest order**
  - Block 1 consists of all treatments with the contrast coefficient equal to -1
  - Block 2 consists of all treatments with the contrast coefficient equal to 1

- Example 1. Block 2^3 Design

factorial effects (contrasts)							
I	A	B	C	AB	AC	BC	ABC
1	-1	-1	-1	1	1	1	-1
1	1	-1	-1	-1	-1	1	1
1	-1	1	-1	-1	1	-1	1
1	1	1	-1	1	-1	-1	-1
1	-1	-1	1	1	-1	-1	1
1	1	-1	1	-1	1	-1	-1
1	-1	1	1	-1	-1	1	-1
1	1	1	1	1	1	1	1

– Defining relation: $b = ABC$:

Block 1: $(- - -)$, $(+ + -)$, $(+ - +)$, $(- + +)$

Block 2: $(+ - -)$, $(- + -)$, $(- - +)$, $(+ + +)$

- Example 2: For 2^4 design with factors: A, B, C, D, the defining contrast (optimal) for blocking factor (b) is

$$b = ABCD$$

- In general, the optimal blocking scheme for 2^k design with two blocks is given by $b = AB \ldots K$, where A, B, ..., K are the factors.

Analyze Unreplicated Block 2^k Experiment

Filtration Experiment (four factors: A, B, C, D):

- Use defining relation: $b = ABCD$
  - If a treatment satisfies $ABCD = -1$, it is allocated to block 1(b_1);
  - If $ABCD = 1$, it is allocated to block 2 (b_2).
- **ASSUME:** all the observations in block 2 will be reduced by 20 because of the poor quality of the second batch of material
  - The true block effect=-20

factor				blocks	
A	B	C	D	$b = ABCD$	response
$-$	$-$	$-$	$-$	1=b_2	45-20=25
$+$	$-$	$-$	$-$	-1=b_1	71
$-$	$+$	$-$	$-$	-1=b_1	48
$+$	$+$	$-$	$-$	1=b_2	65-20=45
$-$	$-$	$+$	$-$	-1=b_1	68
$+$	$-$	$+$	$-$	1=b_2	60-20=40
$-$	$+$	$+$	$-$	1=b_2	80-20=60
$+$	$+$	$+$	$-$	-1=b_1	65
$-$	$-$	$-$	$+$	-1=b_1	43
$+$	$-$	$-$	$+$	1=b_2	100-20=80
$-$	$+$	$-$	$+$	1=b_2	45-20=25
$+$	$+$	$-$	$+$	-1=b_1	104
$-$	$-$	$+$	$+$	1=b_2	75-20=55
$+$	$-$	$+$	$+$	-1=b_1	86
$-$	$+$	$+$	$+$	-1=b_1	70
$+$	$+$	$+$	$+$	1=b_2	96-20=76

SAS File for Block Filtration Experiment

```
data filter;
  do D = -1 to 1 by 2;do C = -1 to 1 by 2;
    do B = -1 to 1 by 2;do A = -1 to 1 by 2;
      input y @@;  output;
    end; end;
  end; end;
  cards;
25 71 48 45 68 40 60 65 43 80 25 104 55 86 70 76
;

data inter;
  set filter;
  AB=A*B; AC=A*C; AD=A*D; BC=B*C; BD=B*D; CD=C*D;
  ABC=AB*C; ABD=AB*D; ACD=AC*D; BCD=BC*D; block=ABC*D;

proc glm data=inter;
  class A B C D AB AC AD BC BD CD ABC ABD ACD BCD block;
  model y=block A B C D AB AC AD BC BD CD ABC ABD ACD BCD; run;
```

```
proc reg outest=effects data=inter;
  model y=A B C D AB AC AD BC BD CD ABC ABD ACD BCD block; run;

data effect2; set effects; drop y intercept _RMSE_;
proc transpose data=effect2 out=effect3;

data effect4; set effect3; effect=col1*2;
proc sort data=effect4; by effect;
proc print data=effect4;

data effect5; set effect4; where  _NAME_^='block';
proc print data=effect5; run;

proc rank data=effect5 normal=blom;
  var effect; ranks neff;

symbol1 v=circle;
proc gplot; plot effect*neff=_NAME_; run; quit;
```

SAS Output: ANOVA Table

Source	DF	Squares	Mean Square	F Value	Pr > F
Model	15	7110.937500	474.062500	.	.
Error		0	0.000000	.	
Co Total	15	7110.937500			

Source	DF	Type I SS	Mean Square	F Value	Pr > F
block	1	1387.562500	1387.562500	.	.
A	1	1870.562500	1870.562500	.	.
B	1	39.062500	39.062500	.	.
C	1	390.062500	390.062500	.	.
D	1	855.562500	855.562500	.	.
AB	1	0.062500	0.062500	.	.
AC	1	1314.062500	1314.062500	.	.
AD	1	1105.562500	1105.562500	.	.
BC	1	22.562500	22.562500	.	.
BD	1	0.562500	0.562500	.	.
CD	1	5.062500	5.062500	.	.
ABC	1	14.062500	14.062500	.	.
ABD	1	68.062500	68.062500	.	.
ACD	1	10.562500	10.562500	.	.
BCD	1	27.562500	27.562500		

- Proportion of variance explained by blocks

$$\frac{1387.5625}{7110.9375} = 19.5\%$$

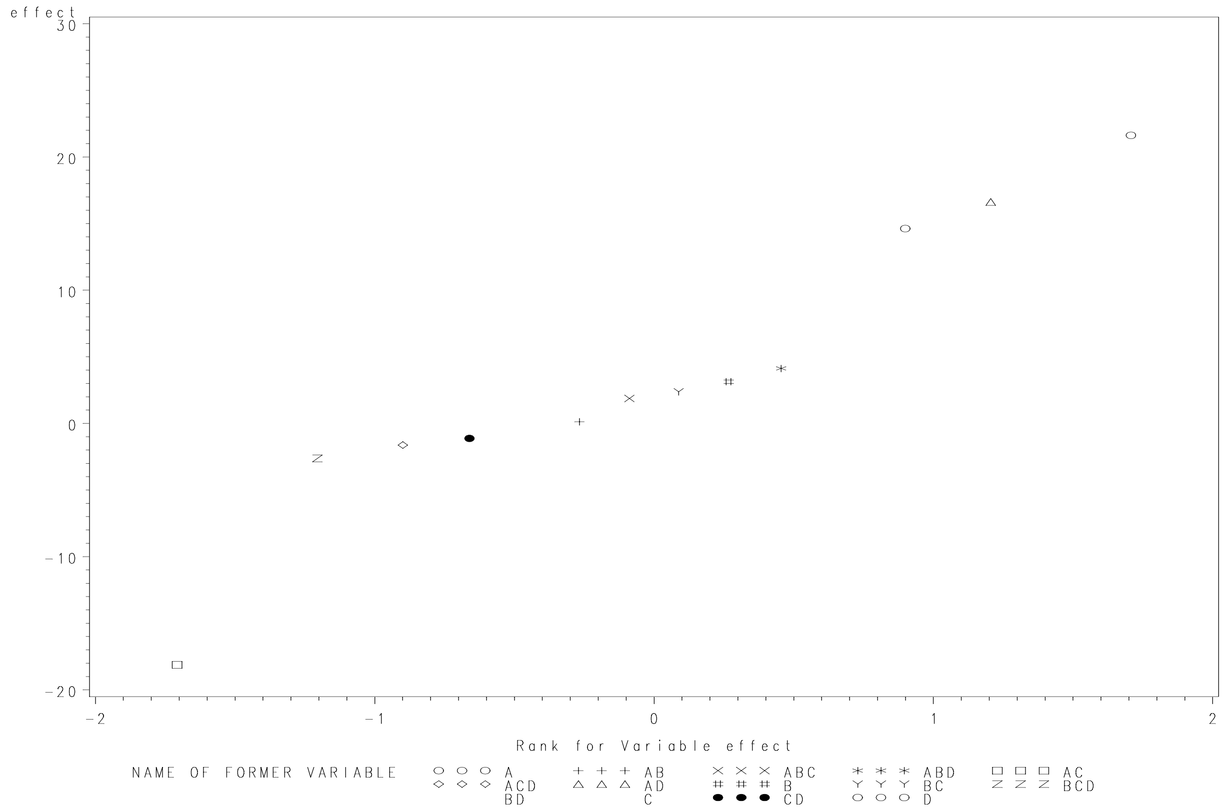
- Similarly proportion of variance can be calculated for other effects.

SAS Output: Factorial Effects and Block Effect

Obs	_NAME_	COL1	effect
1	block	-9.3125	-18.625
2	AC	-9.0625	-18.125
3	BCD	-1.3125	-2.625
4	ACD	-0.8125	-1.625
5	CD	-0.5625	-1.125
6	BD	-0.1875	-0.375
7	AB	0.0625	0.125
8	ABC	0.9375	1.875
9	BC	1.1875	2.375
10	B	1.5625	3.125
11	ABD	2.0625	4.125
12	C	4.9375	9.875
13	D	7.3125	14.625
14	AD	8.3125	16.625
15	A	10.8125	21.625

- Factorial effects are exactly the same as those from the original data (why?)
- Blocking effect: $\bar{y}_{b_2} - \bar{y}_{b_1} = -18.625 = -20 + 1.375$
  - Caused by confounding between b ($= -20$) and $ABCD$ (≈ 1.375).

SAS Output: QQ plot Without Blocking Effect



- Significant effects are:

$$A, C, D, AC, AD$$

2^k Design with Four Blocks

- Need two 2-level blocking factors to generate 4 different blocks.
- Confound each blocking factors with a high order factorial effect.
- The interaction between these two blocking factors matters.
- The interaction will be confounded with another factorial effect.
- Optimal blocking scheme has least confounding severity.
- Example: factors are A, B, C, D and the blocking factors are b_1 and b_2

A	B	C	D	AB	AC	......	CD	ABC	ABD	ACD	BCD	ABCD
−1	−1	−1	−1	1	1		1	−1	−1	−1	−1	1
1	−1	−1	−1	−1	−1		1	1	1	1	−1	−1
−1	1	−1	−1	−1	1		1	1	1	−1	1	−1
1	1	−1	−1	1	−1		1	−1	−1	1	1	1
.	.	.	.	.	.		.	.	.	.	.	.
.	.	.	.	.	.		.	.	.	.	.	.
.	.	.	.	.	.		.	.	.	.	.	.
−1	−1	1	1	1	−1		1	1	1	−1	−1	1
1	1	1	1	−1	1		1	−1	−1	1	−1	−1
−1	−1	1	1	−1	−1		1	−1	−1	−1	1	−1
1	1	1	1	1	1		1	1	1	1	1	1

b1	b2	blocks
−1	−1	1
1	−1	2
−1	1	3
1	1	4

Possible blocking schemes:

- Scheme 1:
  - Defining relations: $b_1 = ABC$, $b_2 = ACD$; induce confounding

$$b_1b_2 = ABC * ACD = A^2BC^2D = BD$$

- Scheme 2:
  - Defining relations: $b_1 = ABCD$, $b_2 = ABC$, induce confounding

$$b_1b_2 = ABCD * ABC = D$$

- Which is better?

2^k Design with 2^p Blocks

- k factors: A,B,...K, and p is usually much less than k.
- p blocking factors: b_1, b_2,...b_p with levels -1 and 1
- Confound blocking factors with p chosen high-order factorial effects, i.e., b_1=effect1, b_2=effect2, etc.(p defining relations)
- These p defining relations induce another $2^p - p - 1$ confounding.
- Treatment combinations with the same values of b_1,...b_p are allocated to the same block. Within each block.
- Each block consists of 2^{k-p} treatment combinations (runs)
- Given k and p, optimal schemes are tabulated, e.g., Montgomery Table 7.9, or Wu & Hamada Appendix 3A