Lec 9: Blocking and Confounding for 2^k Factorial Design

Ying Li

December 2, 2011

2^k factorial design

- Special case of the general factorial design; k factors, all at two levels
- The two levels are usually called low and high (they could be either quantitative or qualitative)
- Very widely used in industrial experimentation

Example

Consider an investigation into the effect of the concentration of the reactant and the amount of the catalyst on the conversion in a chemical process.

- A: reactant concentration, 2 levels
- B: catalyst, 2 levels
- 3 replicates, 12 runs in total

Factor		Treatment	Replicate			
A	B	Combination	I	II	III	Total
$-$	$-$	A low, B low	28	25	27	80
$+$	$-$	A high, B low	36	32	32	100
$-$	$+$	A low, B high	18	19	23	60
$+$	$+$	A high, B high	31	30	29	90

A	B	
$-$	$-$	$(1) = 28 + 25 + 27 = 80$
$+$	$-$	$a = 36 + 32 + 32 = 100$
$-$	$+$	$b = 18 + 19 + 23 = 60$
$+$	$+$	$ab = 31 + 30 + 29 = 90$

$$A = \frac{1}{2n}\{[ab - b] + [a - (1)]\}$$

$$B = \frac{1}{2n}\{[ab - a] + [b - (1)]\}$$

$$AB = \frac{1}{2n}\{[ab - b] - [a - (1)]\}$$

Manual Calculation

- $$A = \frac{1}{2n}\{[ab - b] + [a - (1)]\}$$
- $$Contrast_A = ab + a - b - (1)$$
- $$SS_A = \frac{Contrast_A}{4n}$$

Regression Model

For $2^2 \times 1$ experiment

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_{12} x_1 x_2 + \varepsilon$$

$$\begin{aligned}
(1) &= \beta_0 + \beta_1(-1) + \beta_2(-1) + \beta_{12}(-1)(-1) + \varepsilon_1 \\
a &= \beta_0 + \beta_1(1) + \beta_2(-1) + \beta_{12}(1)(-1) + \varepsilon_2 \\
b &= \beta_0 + \beta_1(-1) + \beta_2(1) + \beta_{12}(-1)(1) + \varepsilon_3 \\
ab &= \beta_0 + \beta_1(1) + \beta_2(1) + \beta_{12}(1)(1) + \varepsilon_4
\end{aligned}$$

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon},\ \mathbf{y} = \begin{bmatrix} (1) \\ a \\ b \\ ab \end{bmatrix}, \mathbf{X} = \begin{bmatrix} 1 & -1 & -1 & 1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & 1 & -1 \\ 1 & 1 & 1 & 1 \end{bmatrix}, \boldsymbol{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \beta_{12} \end{bmatrix}, \boldsymbol{\varepsilon} = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \\ \varepsilon_4 \end{bmatrix}$$

Regression Model

The least square estimates:

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$$

$$= \begin{bmatrix} 4 & 0 & 0 & 0 \\ 0 & 4 & 0 & 0 \\ 0 & 0 & 4 & 0 \\ 0 & 0 & 0 & 4 \end{bmatrix}^{-1} \begin{bmatrix} (1)+a+b+ab \\ a+ab-b-(1) \\ b+ab-a-(1) \\ (1)-a-b+ab \end{bmatrix}$$

$$\begin{bmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \\ \hat{\beta}_2 \\ \hat{\beta}_{12} \end{bmatrix} = \frac{1}{4}\mathbf{I}_4 \begin{bmatrix} (1)+a+b+ab \\ a+ab-b-(1) \\ b+ab-a-(1) \\ (1)-a-b+ab \end{bmatrix} = \begin{bmatrix} \frac{(1)+a+b+ab}{4} \\ \frac{a+ab-b-(1)}{4} \\ \frac{b+ab-a-(1)}{4} \\ \frac{(1)-a-b+ab}{4} \end{bmatrix}$$

The regression coefficient estimates are exactly half of the "usual" effect estimates

Analysis Procedure for a Factorial Design

- Estimate factor effects.
- Formulate model
- Statistical testing (ANOVA).
- Refine the model
- Analyze residuals (graphical)
- Interpret results

Unreplicated 2^k Design

- Lack of replication causes potential problems in statistical testing
  - Replication admits an estimate of pure error (a better phrase is an internal estimate of error)
  - With no replication, fitting the full model results in zero degrees of freedom for error
- Potential solutions to this problem
  - Pooling high-order interactions to estimate error
  - Normal probability plotting of effects

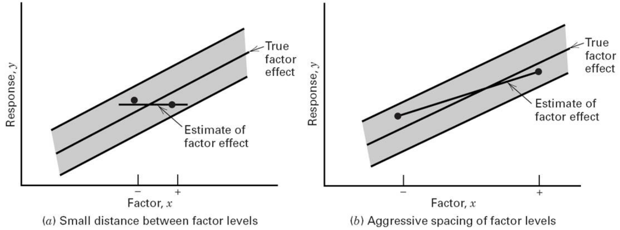


FIGURE 6.9 The impact of the choice of factor levels in an unreplicated design

- If the factors are spaced too closely, it increases the chances that the noise will overwhelm the signal in the data.
- More aggressive spacing is usually best.

Why blocks?

For example, because of

- Different batches of raw material.
- Restrictions caused by equipment and instruments.
- Time.

Factorial design with replicates as blocks

Example

■ **TABLE 7.1**

Chemical Process Experiment in Three Blocks

	Block 1	Block 2	Block 3
	$(1) = 28$	$(1) = 25$	$(1) = 27$
	$a = 36$	$a = 32$	$a = 32$
	$b = 18$	$b = 19$	$b = 23$
	$ab = 31$	$ab = 30$	$ab = 29$
Block totals:	$B_1 = 113$	$B_2 = 106$	$B_3 = 111$

Analysis result

TABLE 7.2
Analysis of Variance for the Chemical Process Experiment in Three Blocks

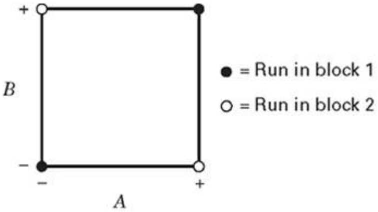
Source of Variation	Sum of Squares	Degrees of Freedom	Mean Square	F_0	P-Value
Blocks	6.50	2	3.25		
A (concentration)	208.33	1	208.33	50.32	0.0004
B (catalyst)	75.00	1	75.00	18.12	0.0053
AB	8.33	1	8.33	2.01	0.2060
Error	24.84	6	4.14		
Total	323.00	11			

A special situation:

- only 1 replicate
- a batch of raw material is only large enough for 2 combinations.

A special situation:

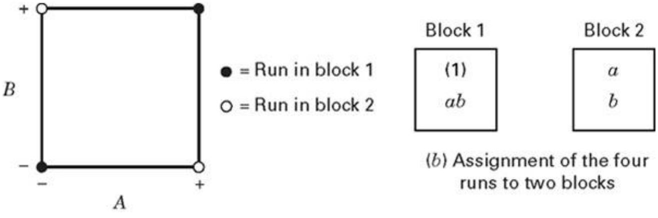
- only 1 replicate
- a batch of raw material is only large enough for 2 combinations.



(a) Geometric view

Block 1	Block 2
(1)	a
ab	b

(b) Assignment of the four runs to two blocks



(a) Geometric view

(b) Assignment of the four runs to two blocks

The main effect of A, B and AB:

$$A = \frac{1}{2}[ab + a - b - (1)]$$

$$B = \frac{1}{2}[ab + b - a - (1)]$$

$$AB = \frac{1}{2}[ab + (1) - a - b]$$

AB is **confounded** with blocks.

Confounding

- a design technique for arranging a complete factorial experiment in blocks.
- the technique causes information about certain treatment effects (usually higher-order interaction) to be **indistinguishable form, or confounded with, blocks**.

■ **TABLE 7.3**

Table of Plus and Minus Signs for the 2^2 Design

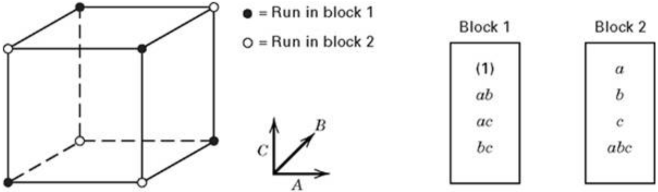
Treatment Combination	Factorial Effect I	A	B	AB	Block
(1)	+	−	−	+	2
a	+	+	−	−	1
b	+	−	+	−	1
ab	+	+	+	+	2

This approach can be used to confound any 2^k design in two blocks.

■ **TABLE 7.4**
Table of Plus and Minus Signs for the 2^3 Design

Treatment Combination	Factorial Effect								
	I	A	B	AB	C	AC	BC	ABC	Block
(1)	+	−	−	+	−	+	+	−	1
a	+	+	−	−	−	−	+	+	2
b	+	−	+	−	−	+	−	+	2
ab	+	+	+	+	−	−	−	−	1
c	+	−	−	+	+	−	−	+	2
ac	+	+	−	−	+	+	−	−	1
bc	+	−	+	−	+	−	+	−	1
abc	+	+	+	+	+	+	+	+	2

The 2^3 design in two blocks with ABC confounded:



(a) Geometric view

(b) Assignment of the eight runs to two blocks

Use defining contrast for constructing the blocks

A defining contrast

$$L = \alpha_1 x_1 + \alpha_2 x_2 + ... + \alpha_k x_k$$

- x_i; the level of ith factor
- α_i is exponent appearing on the ith factor in the effect to to be confounded.

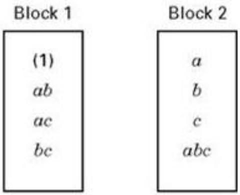
Example

Consider a 2^3 design with ABC confounded with blocks. The defining contrast corresponding to ABC is

$$L = x_1 + x_2 + x_3$$

(1): $L = 0 + 0 + 0 = 0$
ab: $L = 1 + 1 + 0 = 2$
ac: $L = 1 + 0 + 1 = 2$
bc: $L = 0 + 1 + 1 = 2$
a: $L = 1 + 0 + 0 = 1$
b: $L = 0 + 1 + 0 = 1$
c: $L = 0 + 0 + 1 = 1$
abc: $L = 1 + 1 + 1 = 3$

Use the principal block for constructing the blocks



(b) Assignment of the eight runs to two blocks

- The block containing (1) is called the **principal block**.
- Any element in the principal block may be generated by multiplying two other elements in the principal block modulus 2.
- Treatment combinations in the other block may be generated by multiplying one element in the new block by each element in the principal block modulus 2.

Four blocks in a 2^5

Four blocks in a 2^5

- Select two defining interactions
- Say ADE and BCE, the two defining contrasts are

$$L_1 = x_1 + x_4 + x_5$$

$$L_2 = x_2 + x_3 + x_5$$

So $(L_1, L_2) = (0,0), (0,1), (1,0), \text{or} (1,1)$

Block 1	Block 2	Block 3	Block 4
$L_1 = 0$	$L_1 = 1$	$L_1 = 0$	$L_1 = 1$
$L_2 = 0$	$L_2 = 0$	$L_2 = 1$	$L_2 = 1$
(1) abe	a be	b $abce$	e $abcde$
ad ace	d $abde$	abd ae	ade bd
bc cde	abc ce	c $bcde$	bce ac
$abcd$ bde	bcd $acde$	acd de	ab cd

Generalized interaction

It is defined as the product of ADE and BCE modulus 2.

$$(ADE)(BCE) = ABCDE^2 = ABCD$$

ADE, BCE and ABCD are confounded with blocks in this design

Question

Which alternative is better:

1. ADE and BCE or
2. ABCDE and ABD??

■ TABLE 7.9

Suggested Blocking Arrangements for the 2^k Factorial Design

Number of Factors, k	Number of Blocks, 2^p	Block Size, 2^{k-p}	Effects Chosen to Generate the Blocks	Interactions Confounded with Blocks
3	2	4	*ABC*	*ABC*
	4	2	*AB, AC*	*AB, AC, BC*
4	2	8	*ABCD*	*ABCD*
	4	4	*ABC, ACD*	*ABC, ACD, BD*
	8	2	*AB, BC, CD*	*AB, BC, CD, AC, BD, AD, ABCD*
5	2	16	*ABCDE*	*ABCDE*
	4	8	*ABC, CDE*	*ABC, CDE, ABDE*
	8	4	*ABE, BCE, CDE*	*ABE, BCE, CDE, AC, ABCD, BD, ADE*
	16	2	*AB, AC, CD, DE*	All two- and four-factor interactions (15 effects)
6	2	32	*ABCDEF*	*ABCDEF*
	4	16	*ABCF, CDEF*	*ABCF, CDEF, ABDE*
	8	8	*ABEF, ABCD, ACE*	*ABEF, ABCD, ACE, BCF, BDE, CDEF, ADF*
	16	4	*ABF, ACF, BDF, DEF*	*ABF, ACF, BDF, DEF, BC, ABCD, ABDE, AD, ACDE, CE, CDF, BCDEF, ABCEF, AEF, BE*
	32	2	*AB, BC, CD, DE, EF*	All two-, four-, and six-factor interactions (31 effects)
7	2	64	*ABCDEFG*	*ABCDEFG*
	4	32	*ABCFG, CDEFG*	*ABCFG, CDEFG, ABDE*
	8	16	*ABC, DEF, AFG*	*ABC, DEF, AFG, ABCDEF, BCFG, ADEG, BCDEG*
	16	8	*ABCD, EFG, CDE, ADG*	*ABCD, EFG, CDE, ADG, ABCDEFG, ABE, BCG, CDFG, ADEF, ACEG, ABFG, BCEF, BDEG, ACF, BDF*
	32	4	*ABG, BCG, CDG, DEG, EFG*	*ABG, BCG, CDG, DEG, EFG, AC, BD, CE, DF, AE, BF, ABCD, ABDE, ABEF, BCDE, BCEF, CDEF, ABCDEFG, ADG, ACDEG, ACEFG, ABDFG, ABCEG, BEG, BDEFG, CFG, ADEF, ACDF, ABCF, AFG, BCDFG*
	64	2	*AB, BC, CD, DE, EF, FG*	All two-, four-, and six-factor interactions (63 effects)

Example

A chemical product is produced in a pressure vessel. A factorial experiment is carried out in the pilot plant to study the factors thought to influence the filtration rate of this product.
The four factors are all present at two levels. They are:

- A: temperature
- B: pressure
- C: concentration of formaldehyde
- D: stirring rate

■ **TABLE 6.11**

Contrast Constants for the 2^4 Design

	A	B	AB	C	AC	BC	ABC	D	AD	BD	ABD	CD	ACD	BCD	ABCD
(1)	−	−	+	−	+	+	−	−	+	+	−	+	−	−	+
a	+	−	−	−	−	+	+	−	−	+	+	+	+	−	−
b	−	+	−	−	+	−	+	−	+	−	+	+	−	+	−
ab	+	+	+	−	−	−	−	−	−	−	−	+	+	+	+
c	−	−	+	+	−	−	+	−	+	+	−	−	+	+	−
ac	+	−	−	+	+	−	−	−	−	+	+	−	−	+	+
bc	−	+	−	+	−	+	−	−	+	−	+	−	+	−	+
abc	+	+	+	+	+	+	+	−	−	−	−	−	−	−	−
d	−	−	+	−	+	+	−	+	−	−	+	−	+	+	−
ad	+	−	−	−	−	+	+	+	+	−	−	−	−	+	+
bd	−	+	−	−	+	−	+	+	−	+	−	−	+	−	+
abd	+	+	+	−	−	−	−	+	+	+	+	−	−	−	−
cd	−	−	+	+	−	−	+	+	−	−	+	+	−	−	+
acd	+	−	−	+	+	−	−	+	+	−	−	+	+	−	−
bcd	−	+	−	+	−	+	−	+	−	+	−	+	−	+	−
abcd	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+

■ **TABLE 6.10**

Pilot Plant Filtration Rate Experiment

Run Number	Factor A	B	C	D	Run Label	Filtration Rate (gal/h)
1	−	−	−	−	(1)	45
2	+	−	−	−	*a*	71
3	−	+	−	−	*b*	48
4	+	+	−	−	*ab*	65
5	−	−	+	−	*c*	68
6	+	−	+	−	*ac*	60
7	−	+	+	−	*bc*	80
8	+	+	+	−	*abc*	65
9	−	−	−	+	*d*	43
10	+	−	−	+	*ad*	100
11	−	+	−	+	*bd*	45
12	+	+	−	+	*abd*	104
13	−	−	+	+	*cd*	75
14	+	−	+	+	*acd*	86
15	−	+	+	+	*bcd*	70
16	+	+	+	+	*abcd*	96

■ **TABLE 6.10**
Pilot Plant Filtration Rate Experiment

Run Number	Run Label	Filtration Rate (gal/h)
1	(1)	45
2	*a*	71
3	*b*	48
4	*ab*	65
5	*c*	68
6	*ac*	60
7	*bc*	80
8	*abc*	65
9	*d*	43
10	*ad*	100
11	*bd*	45
12	*abd*	104
13	*cd*	75
14	*acd*	86
15	*bcd*	70
16	*abcd*	96

Block 1	Block 2
$(1) = 25$	$a = 71$
$ab = 45$	$b = 48$
$ac = 40$	$c = 68$
$bc = 60$	$d = 43$
$ad = 80$	$abc = 65$
$bd = 25$	$bcd = 70$
$cd = 55$	$acd = 86$
$abcd = 76$	$abd = 104$

(*b*) Assignment of the 16 runs to two blocks

■ TABLE 7.6

Effect Estimates for the Blocked 2^4 Design in Example 7.2

Model Term	Regression Coefficient	Effect Estimate	Sum of Squares	Percent Contribution
A	10.81	21.625	1870.5625	26.30
B	1.56	3.125	39.0625	0.55
C	4.94	9.875	390.0625	5.49
D	7.31	14.625	855.5625	12.03
AB	0.062	0.125	0.0625	<0.01
AC	−9.06	−18.125	1314.0625	18.48
AD	8.31	16.625	1105.5625	15.55
BC	1.19	2.375	22.5625	0.32
BD	−0.19	−0.375	0.5625	<0.01
CD	−0.56	−1.125	5.0625	0.07
ABC	0.94	1.875	14.0625	0.20
ABD	2.06	4.125	68.0625	0.96
ACD	−0.81	−1.625	10.5625	0.15
BCD	−1.31	−2.625	27.5625	0.39
Block ($ABCD$)		−18.625	1387.5625	19.51

■ **TABLE 6.12**

Factor Effect Estimates and Sums of Squares for the 2^4 Factorial in Example 6.2

Model Term	Effect Estimate	Sum of Squares	Percent Contribution
A	21.625	1870.56	32.6397
B	3.125	39.0625	0.681608
C	9.875	390.062	6.80626
D	14.625	855.563	14.9288
AB	0.125	0.0625	0.00109057
AC	−18.125	1314.06	22.9293
AD	16.625	1105.56	19.2911
BC	2.375	22.5625	0.393696
BD	−0.375	0.5625	0.00981515
CD	−1.125	5.0625	0.0883363
ABC	1.875	14.0625	0.245379
ABD	4.125	68.0625	1.18763
ACD	−1.625	10.5625	0.184307
BCD	−2.625	27.5625	0.480942
$ABCD$	1.375	7.5625	0.131959

Partial confounding

Different confounding in different replicates.

Instead of this...

Replicate I		Replicate II		Replicate III		Replicate IV	
Block 1	Block 2	Block 1	Block 2	Block 1	Block 2	Block 1	Block 2
(1)	abc	(1)	abc	(1)	abc	(1)	abc
ac	a	ac	a	ac	a	ac	a
ab	b	ab	b	ab	b	ab	b
bc	c	bc	c	bc	c	bc	c

… we can do this

Replicate I *ABC* Confounded		Replicate II *AB* Confounded		Replicate III *BC* Confounded		Replicate IV *AC* Confounded	
(1)	*a*	(1)	*a*	(1)	*b*	(1)	*a*
ab	*b*	*c*	*b*	*a*	*c*	*b*	*c*
ac	*c*	*ab*	*ac*	*bc*	*ab*	*ac*	*ab*
bc	*abc*	*abc*	*bc*	*abc*	*ac*	*abc*	*bc*

Recommended problems

- 7.2
- 7.18
- 7.19
- 7.20