Topic 15: blocking and confounding 2^K factorial design

Montgomery: chapter 7

Prof. Lingling An
University of Arizona

2^4 factorial design

A	B	C	D	rep1	rep2	rep3
-	-	-	-			
+	-	-	-			
-	+	-	-			
+	+	-	-			
-	-	+	-			
+	-	+	-			
-	+	+	-			
+	+	+	-			
-	-	-	+			
+	-	-	+			
-	+	-	+			
+	+	-	+			
-	-	+	+			
+	-	+	+			
-	+	+	+			
+	+	+	+			

Blocking a replicated 2^4 factorial design

A	B	C	D	Day 1	Day 2	Day 3
-	-	-	-			
+	-	-	-			
-	+	-	-			
+	+	-	-			
-	-	+	-			
+	-	+	-			
-	+	+	-			
+	+	+	-			
-	-	-	+			
+	-	-	+			
-	+	-	+			
+	+	-	+			
-	-	+	+			
+	-	+	+			
-	+	+	+			
+	+	+	+			

Unreplicated 2^4 factorial design

A	B	C	D	response
-	-	-	-	
+	-	-	-	
-	+	-	-	
+	+	-	-	
-	-	+	-	
+	-	+	-	
-	+	+	-	
+	+	+	-	
-	-	-	+	
+	-	-	+	
-	+	-	+	
+	+	-	+	
-	-	+	+	
+	-	+	+	
-	+	+	+	
+	+	+	+	

confounding 2^4 factorial design in two blocks

A	B	C	D	response
-	-	-	-	
+	-	-	-	
-	+	-	-	
+	+	-	-	
-	-	+	-	
+	-	+	-	
-	+	+	-	
+	+	+	-	
-	-	-	+	
+	-	-	+	
-	+	-	+	
+	+	-	+	
-	-	+	+	
+	-	+	+	
-	+	+	+	
+	+	+	+	

Recall: Blocking in Factorial Design - Example

Battery Life Experiment:

An engineer is studying the effective lifetime of some battery. Two factors, plate material and temperature, are involved. There are three types of plate materials (1, 2, 3) and three temperature levels (15, 70, 125). Four batteries are tested at each combination of plate material and temperature, and all 36 tests are run in random order. The experiment and the resulting observed battery life data are given below.

material	temperature 15	70	125
1	130,155,74,180	34,40,80,75	20,70,82,58
2	150,188,159,126	136,122,106,115	25,70,58,45
3	138,110,168,160	174,120,150,139	96,104,82,60

If we assume further that four operators (1,2,3,4) were hired to conduct the experiment. It is known that different operators can cause systematic difference in battery lifetime. Hence operators should be treated as blocks

The blocking scheme is every operator conduct a single replicate of the full factorial design

For each treatment (treatment combination), the observations were in the order of the operators 1, 2, 3, and 4.

This is a blocked factorial design

Statistical Model for Blocked Factorial Experiment

$$y_{ijk} = \mu + \tau_i + \beta_j + (\tau\beta)_{ij} + \delta_k + \epsilon_{ijk}$$

$i = 1, 2, \ldots, a$, $j = 1, 2, \ldots, b$ and $k = 1, 2, \ldots, n$, δ_k is the effect of the kth block.

- randomization restriction is imposed. (complete block factorial design).
- interactions between blocks and treatment effects are assumed to be negligible.
- The previous ANOVA table for the experiment should be modified as follows:

  Add: Block Sum of Square

$$SS_{Blocks} = \frac{1}{ab}\sum_k y_{..k}^2 - \frac{y_{...}^2}{abn} \text{ D.F. } n-1$$

  Modify: Error Sum of Squares:

$$\text{(new)}SS_E = \text{(old)}SS_E - SS_{\text{Blocks}} \text{ D.F. } (ab-1)(n-1)$$

- other inferences should be modified accordingly.

- Now consider the **unreplicated** case
- Clearly the previous discussion does not apply, since there is only one replicate
- To illustrate, consider the situation of Example 6.2, the pilot plant filtration rate experiment
- This is a 2^4, n = 1 replicate

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

Use plot to show significance !

How about … ?

- 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?

A	B	C	D	response
-	-	-	-	
+	-	-	-	
-	+	-	-	
+	+	-	-	
-	-	+	-	
+	-	+	-	
-	+	+	-	
+	+	+	-	
-	-	-	+	
+	-	-	+	
-	+	-	+	
+	+	-	+	
-	-	+	+	
+	-	+	+	
-	+	+	+	
+	+	+	+	

2^2 Design with Two Blocks (note: not 2 block factors!)

Suppose there are two factors (A, B) each with 2 levels, and two blocks (b_1, b_2) each containing 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:

$$\text{block effect: } b = \bar{y}_{b_2} - \bar{y}_{b_1} = \frac{1}{2}(-y_{--} + y_{+-} - y_{-+} + y_{++})$$

$$\text{main effect: } A = \frac{1}{2}(-y_{--} + y_{+-} - y_{-+} + y_{++})$$

A and b are not distinguishable, or confounded.

Scheme 3:

$$\text{block effect: } b = \bar{y}_{b_2} - \bar{y}_{b_1} = \frac{1}{2}(y_{--} - y_{+-} - y_{-+} + y_{++})$$

$$\text{interaction: } AB = \frac{1}{2}(y_{--} - y_{+-} - y_{-+} + y_{++})$$

AB and b become indistinguishable, or confounded.

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)?

2^k Design with Two Blocks via Confounding

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: $b = ABCD$, i.e., 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 that, all the observations in block 2 will be reduced by 20 because of the poor quality of the second batch of material, i.e. 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

```
goption colors=(none);
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;
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;
```

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	.	.

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

Comparison with original data result

factor				
A	*B*	*C*	*D*	filtration
−	−	−	−	45
+	−	−	−	71
−	+	−	−	48
+	+	−	−	65
−	−	+	−	68
+	−	+	−	60
−	+	+	−	80
+	+	+	−	65
−	−	−	+	43
+	−	−	+	100
−	+	−	+	45
+	+	−	+	104
−	−	+	+	75
+	−	+	+	86
−	+	+	+	70
+	+	+	+	96

Obs	_NAME_	COL1	effect
1	AC	-9.0625	-18.125
2	BCD	-1.3125	-2.625
3	ACD	-0.8125	-1.625
4	CD	-0.5625	-1.125
5	BD	-0.1875	-0.375
6	AB	0.0625	0.125
7	ABCD	0.6875	1.375
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

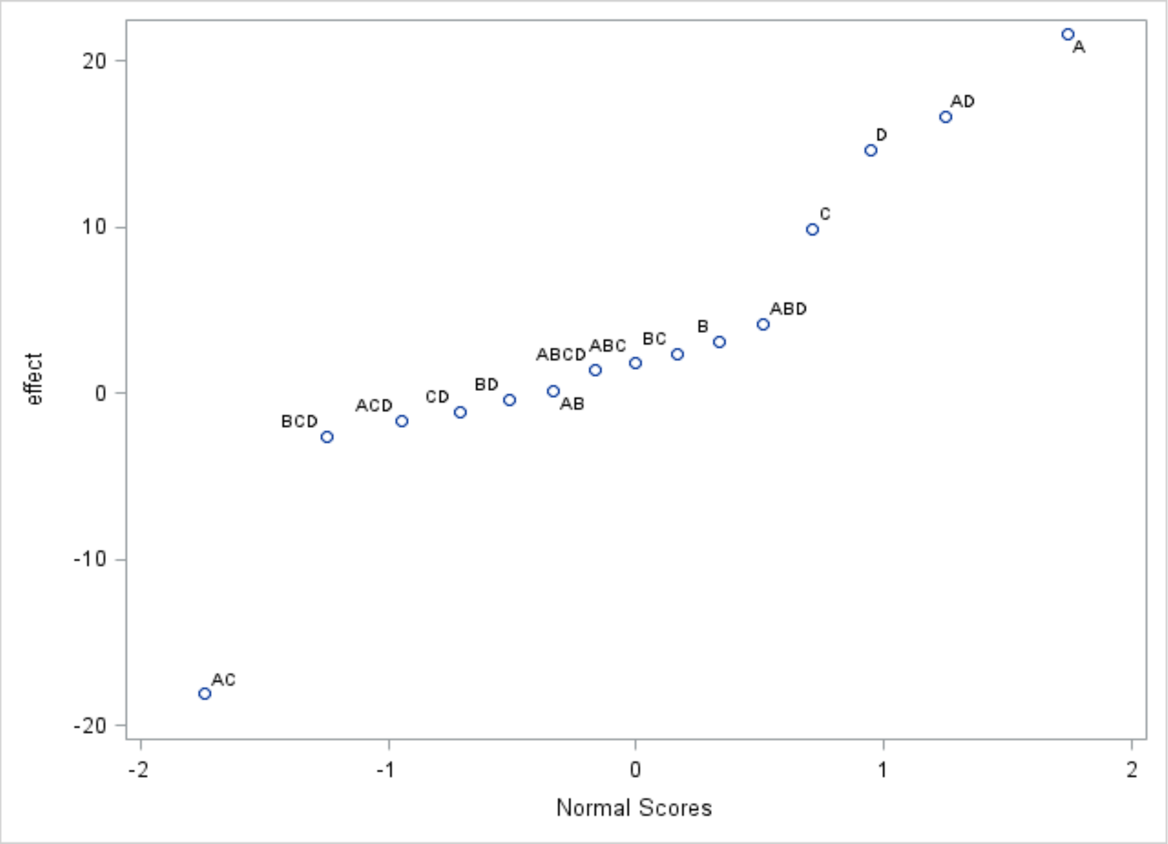
Factorial effects are exactly the same as those from the original data (why?)

blocking effect: -18.625=$\bar{y}_{b_2} - \bar{y}_{b_1}$, is in fact

$$-20(\text{true blocking effect}) + 1.375(\text{some interaction of } ABCD)$$

This is caused by confounding between b and $ABCD$.

- SAS output: QQ plot without blocking effect



Significant effects: A, C, D, AC, AD

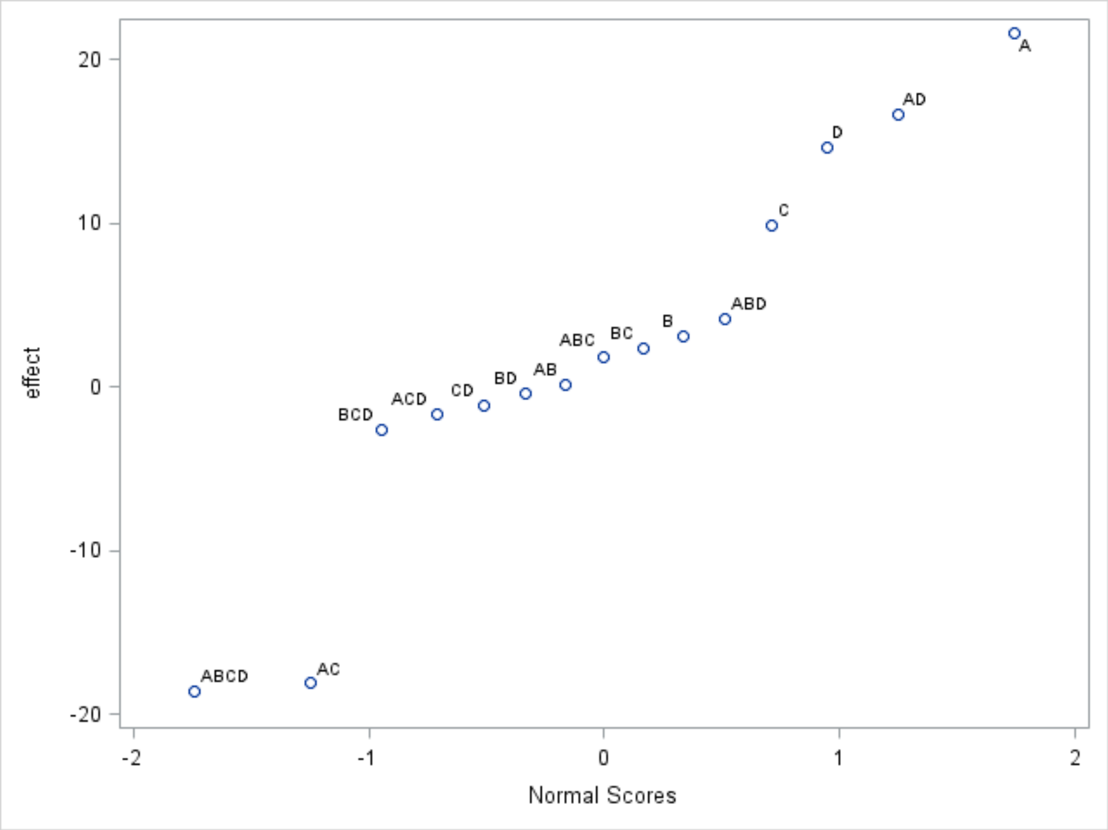
Another illustration of the importance of blocking

■ **TABLE 7.8**
The Modified Data from Example 7.2

Run Order	Std. Order	Factor *A*: Temperature	Factor *B*: Pressure	Factor *C*: Concentration	Factor *D*: Stirring Rate	Response Filtration Rate
8	1	−1	−1	−1	−1	25
11	2	1	−1	−1	−1	71
1	3	−1	1	−1	−1	28
3	4	1	1	−1	−1	45
9	5	−1	−1	1	−1	68
12	6	1	−1	1	−1	60
2	7	−1	1	1	−1	60
13	8	1	1	1	−1	65
7	9	−1	−1	−1	1	23
6	10	1	−1	−1	1	80
16	11	−1	1	−1	1	45
5	12	1	1	−1	1	84
14	13	−1	−1	1	1	75
15	14	1	−1	1	1	86
10	15	−1	1	1	1	70
4	16	1	1	1	1	76

Now the first eight runs (in run order) have filtration rate reduced by 20 units

If ignore the blocking effect in the model …



The interpretation is harder;

Failing to block when we should have causes problems in interpretation the result of an experiment and can mask the presence of real factor effects.

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.

In another word, for **four** blocks, select **two** effects to confound, automatically confounding a **third** effect.

Optimal blocking scheme has least confounding severity.

2^4 design with four blocks: factors are A, B, C, D and the blocking factors are $b1$ and $b2$

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: $b1 = ABC$, $b2 = ACD$; induce confounding

$$b1b2 = ABC * ACD = A^2BC^2D = BD$$

Scheme 2:

Defining relations: $b1 = ABCD$, $b2 = ABC$, induce confounding

$$b1b2 = ABCD * ABC = D$$

Which is better?

■ 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)

General advice about blocking

- When in doubt, block
- Block out the nuisance variables you know about, randomize as much as possible and rely on randomization to help balance out unknown nuisance effects
- Measure the nuisance factors you know about but can't control (ANCOVA)
- It may be a good idea to conduct the experiment in blocks even if there isn't an obvious nuisance factor, just to protect against the loss of data or situations where the complete experiment can't be finished.

Last slide

- Read Sections: 7.1-7.7