Stat 571B Experimental Design

Topic 16: Intro to fractional 2^k factorial design

Montgomery: chapter 8

Prof. Lingling An
University of Arizona

2^4 factorial design

A	B	C	D	rep1	rep2	rep3	rep4
-	-	-	-				
+	-	-	-				
-	+	-	-				
+	+	-	-				
-	-	+	-				
+	-	+	-				
-	+	+	-				
+	+	+	-				
-	-	-	+				
+	-	-	+				
-	+	-	+				
+	+	-	+				
-	-	+	+				
+	-	+	+				
-	+	+	+				
+	+	+	+				

Blocking a replicated 2^4 factorial design

A	B	C	D	Day 1	Day 2	Day 3	Day 4
-	-	-	-				
+	-	-	-				
-	+	-	-				
+	+	-	-				
-	-	+	-				
+	-	+	-				
-	+	+	-				
+	+	+	-				
-	-	-	+				
+	-	-	+				
-	+	-	+				
+	+	-	+				
-	-	+	+				
+	-	+	+				
-	+	+	+				
+	+	+	+				

Unreplicated 2^4 factorial design

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

confounding 2^4 factorial design in two blocks

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

Half of 2^4 factorial design

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

Half of 2^4 factorial design

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

Fundamental Principles Regarding Factorial Effects

Suppose there are k factors (A,B,...,J,K) in an experiment. All possible factorial effects include

effects of order 1: A, B, ..., K (main effects)

effects of order 2: AB, AC, ...,JK (2-factor interactions)

....

- Hierarchical Ordering principle
  - Lower order effects are more likely to be important than higher order effects.
  - Effects of the same order are equally likely to be important
- Effect Sparsity Principle (Pareto principle)
  - The number of relatively important effects in a factorial experiment is small
- Effect Heredity Principle
  - In order for an interaction to be significant, at least one of its parent factors should be significant.

- May not have sources (time,money,etc) for full factorial design
- Number of runs required for full factorial grows quickly
  - Consider 2^k design
  - If $k = 7 \rightarrow$ 128 runs required
  - Can estimate 127 effects
  - Only 7 df for main effects, 21 for 2-factor interactions
  - the remaining 99 df are for interactions of order ≥ 3
- Often only lower order effects are important
- Full factorial design may not be necessary according to
  - Hierarchical ordering principle
  - Effect Sparsity Principle
- A fraction of the full factorial design (i.e. a subset of all possible level combinations) is sufficient.

Fractional Factorial Design

Example 1

- Suppose you were designing a new car
- Wanted to consider the following nine factors each with 2 levels
  - 1. Engine Size; 2. Number of cylinders; 3. Drag; 4. Weight; 5. Automatic vs Manual; 6. Shape; 7. Tires; 8. Suspension; 9. Gas Tank Size;
- Only have resources for conduct $2^6 = 64$ runs
  - If you drop three factors for a 2^6 full factorial design, those factor and their interactions with other factors cannot be investigated.
  - Want investigate all nine factors in the experiment
  - A fraction of 2^9 full factorial design will be used.
  - Confounding (aliasing) will happen because using a subset

How to choose (or construct) the fraction?

Example 2

Filtration rate experiment:

Recall that there are four factors in the experiment(A, B, C and D), each of 2 levels. Suppose the available resource is enough for conducting 8 runs. 2^4 full factorial design consists of all the 16 level combinations of the four factors. We need to choose half of them.

The chosen half is called 2^{4-1} fractional factorial design.

Which half we should select (construct)?

factor			
A	B	C	D
$-$	$-$	$-$	$-$
$+$	$-$	$-$	$-$
$-$	$+$	$-$	$-$
$+$	$+$	$-$	$-$
$-$	$-$	$+$	$-$
$+$	$-$	$+$	$-$
$-$	$+$	$+$	$-$
$+$	$+$	$+$	$-$
$-$	$-$	$-$	$+$
$+$	$-$	$-$	$+$
$-$	$+$	$-$	$+$
$+$	$+$	$-$	$+$
$-$	$-$	$+$	$+$
$+$	$-$	$+$	$+$
$-$	$+$	$+$	$+$
$+$	$+$	$+$	$+$

2^{4-1} Fractional Factorial Design

- the number of factors: $k = 4$
- the fraction index: $p = 1$
- the number of runs (level combinations): $N = \frac{2^4}{2^1} = 8$
- Construct 2^{4-1} designs via "confounding" (**aliasing**)
  - select 3 factors (e.g. A, B, C) to form a 2^3 full factorial (basic design)
  - confound (**alias**) D with a high order interaction of A, B and C. For example,

$$D = ABC$$

factorial effects (contrasts)							
I	A	B	C	AB	AC	BC	ABC=D
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

- Therefore, the chosen fraction includes the following 8 level combinations:
- $(-,-,-,-)$, $(+,-,-,+)$, $(-,+,-,+)$, $(+,+,-,-)$, $(-,-,+,+)$, $(+,-,+,-)$, $(-,+,+,-)$, $(+,+,+,+)$
- Note: 1 corresponds to $+$ and -1 corresponds to $-$.

Verify:

1. the chosen level combinations form a half of the 2^4 design.

2. the product of columns A, B, C and D equals 1, i.e.,

$$I = ABCD$$

which is called the **defining relation**, or $ABCD$ is called a **defining word** (contrast).

Aliasing in 2^{4-1} Design

For four factors A, B, C and D, there are $2^4 - 1$ effects: A, B, C, D, AB, AC, AD, BC, BD, CD, ABC, ABD, ACD, BCD, $ABCD$

Response	I	A	B	C	D	AB	..	CD	ABC	BCD	...	ABCD
y_1	1	-1	-1	-1	-1	1	..	1	-1	-1	...	1
y_2	1	1	-1	-1	1	-1	..	-1	1	1	...	1
y_3	1	-1	1	-1	1	-1	..	-1	1	-1	...	1
y_4	1	1	1	-1	-1	1	..	1	-1	1	...	1
y_5	1	-1	-1	1	1	1	..	1	1	-1	...	1
y_6	1	1	-1	1	-1	-1	..	-1	-1	1	...	1
y_7	1	-1	1	1	-1	-1	..	-1	-1	-1	...	1
y_8	1	1	1	1	1	1	..	1	1	1	...	1

Contrasts for main effects by converting $-$ to -1 and $+$ to 1; contrasts for other effects obtained by multiplication.

$$A = \bar{y}_{A+} - \bar{y}_{A-} = \frac{1}{4}(-y_1 + y_2 - y_3 + y_4 - y_5 + y_6 - y_7 + y_8)$$

$$BCD = \frac{1}{4}(-y_1 + y_2 - y_3 + y_4 - y_5 + y_6 - y_7 + y_8)$$

A, BCD are aliases or aliased. The contrast is for A+BCD. A and BCD are not distinguishable.

$$AB = \bar{y}_{AB+} - \bar{y}_{AB-} = \frac{1}{4}(y_1 - y_2 - y_3 + y_4 + y_5 - y_6 - y_7 + y_8)$$
$$CD = \bar{y}_{CD+} - \bar{y}_{CD-} = \frac{1}{4}(y_1 - y_2 - y_3 + y_4 + y_5 - y_6 - y_7 + y_8)$$

AB, CD are aliases or aliased. The contrast is for AB+CD. AB and CD are not distinguishable.

There are other 5 pairs. They are caused by the defining relation

$$I = ABCD,$$

that is, I (the intercept) and 4-factor interaction $ABCD$ are aliased.

Alias Structure for 2^{4-1} with $I = ABCD$ (denoted by d_1)

- Alias Structure:

  $I = ABCD$

  $A = A * I = A * ABCD = BCD$

  $B = ... = ACD$

  $C = ... = ABD$

  $D = ... = ABC$

  $AB = AB * I = AB * ABCD = CD$

  $AC = ... = BD$

  $AD = ... = BC$

  all 16 factorial effects for A, B, C and D are partitioned into 8 groups each with 2 aliased effects.

A Different 2^{4-1} Fractional Factorial Design

- the defining relation $I = ABD$ generates another 2^{4-1} fractional factorial design, denoted by d_2. Its alias structure is given below.

  $I = ABD$

  $A = BD$

  $B = AD$

  $C = ABCD$

  $D = AB$

  $ABC = CD$

  $ACD = BC$

  $BCD = AC$

- Recall d_1 is defined by $I = ABCD$. Comparing d_1 and d_2, which one we should choose or which one is better?

  1. **Length** of a defining word is defined to be the number of the involved factors.

  2. **Resolution** of a fractioanl factorial design is defined to be the minimum length of the defining words, usually denoted by Roman numbers, III, IV, V, etc...

Resolution and Maximum Resolution Criterion

- d_1: $I = ABCD$ is a resolution IV design denoted by 2^{4-1}_{IV}.
- d_2: $I = ABD$ is a resolution III design denoted by 2^{4-1}_{III}.
- If a design is of resolution R, then none of the i-factor interactions is aliased with any other interaction of order less than $R - i$.

  d_1: main effects are not aliased with other main effects and 2-factor interactions

  d_2: main effects are not aliased with main effects
- d_1 is better, because d_1 has higher resolution than d_2. In fact, d_1 is optimal among all the possible fractional factorial 2^{4-1} designs
- **Maximum Resolution Criterion**

  fractional factorial design with maximum resolution is optimal

Analysis for 2^{4-1} Design: Filtration Experiment

Recall that the filtration rate experiment was originally a 2^4 full factorial experiment. We pretend that only half of the combinations were run. The chosen half is defined by $I = ABCD$. So it is now a 2^{4-1} design. We keep the original responses.

basic design				
A	B	C	$D = ABC$	filtration rate
$-$	$-$	$-$	$-$	45
$+$	$-$	$-$	$+$	100
$-$	$+$	$-$	$+$	45
$+$	$+$	$-$	$-$	65
$-$	$-$	$+$	$+$	75
$+$	$-$	$+$	$-$	60
$-$	$+$	$+$	$-$	80
$+$	$+$	$+$	$+$	96

Let $\mathcal{L}_{\text{effect}}$ denote the estimate of effect (based on the corresponding contrast). Because of aliasing,

$$\mathcal{L}_I \to I + ABCD$$

$$\mathcal{L}_A \to A + BCD$$

$$\mathcal{L}_B \to B + ACD$$

$$\mathcal{L}_C \to C + ABD$$

$$\mathcal{L}_D \to D + ABC$$

$$\mathcal{L}_{AB} \to AB + CD$$

$$\mathcal{L}_{AC} \to AC + BD$$

$$\mathcal{L}_{AD} \to AD + BC$$

SAS file for 2^{4-1} Filtration Experiment

```
goption colors=(none);
data filter;
 do C = -1 to 1 by 2;
 do B = -1 to 1 by 2;do A = -1 to 1 by 2; D=A*B*C;
 input y @@;  output; end; end; end;
 datalines;
 45 100 45 65 75 60 80 96
 ;

data inter;        /* Define Interaction Terms */
set filter;
AB=A*B; AC=A*C; AD=A*D;

proc glm data=inter;    /* GLM Proc to Obtain Effects */
class A B C D AB AC AD;
model y=A B C D AB AC AD;
estimate 'A' A -1 1; estimate 'B' B -1 1; estimate 'C' C -1 1;
estimate 'D' D -1 1; estimate 'AB' AB -1 1; estimate 'AC' AC -1 1;
estimate 'AD' AD -1 1;  run;
```

```
proc reg outest=effects data=inter; /* REG Proc to Obtain Effects */
model y=A B C D AB AC AD;

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;
proc rank data=effect4 normal=blom;
var effect; ranks neff;

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

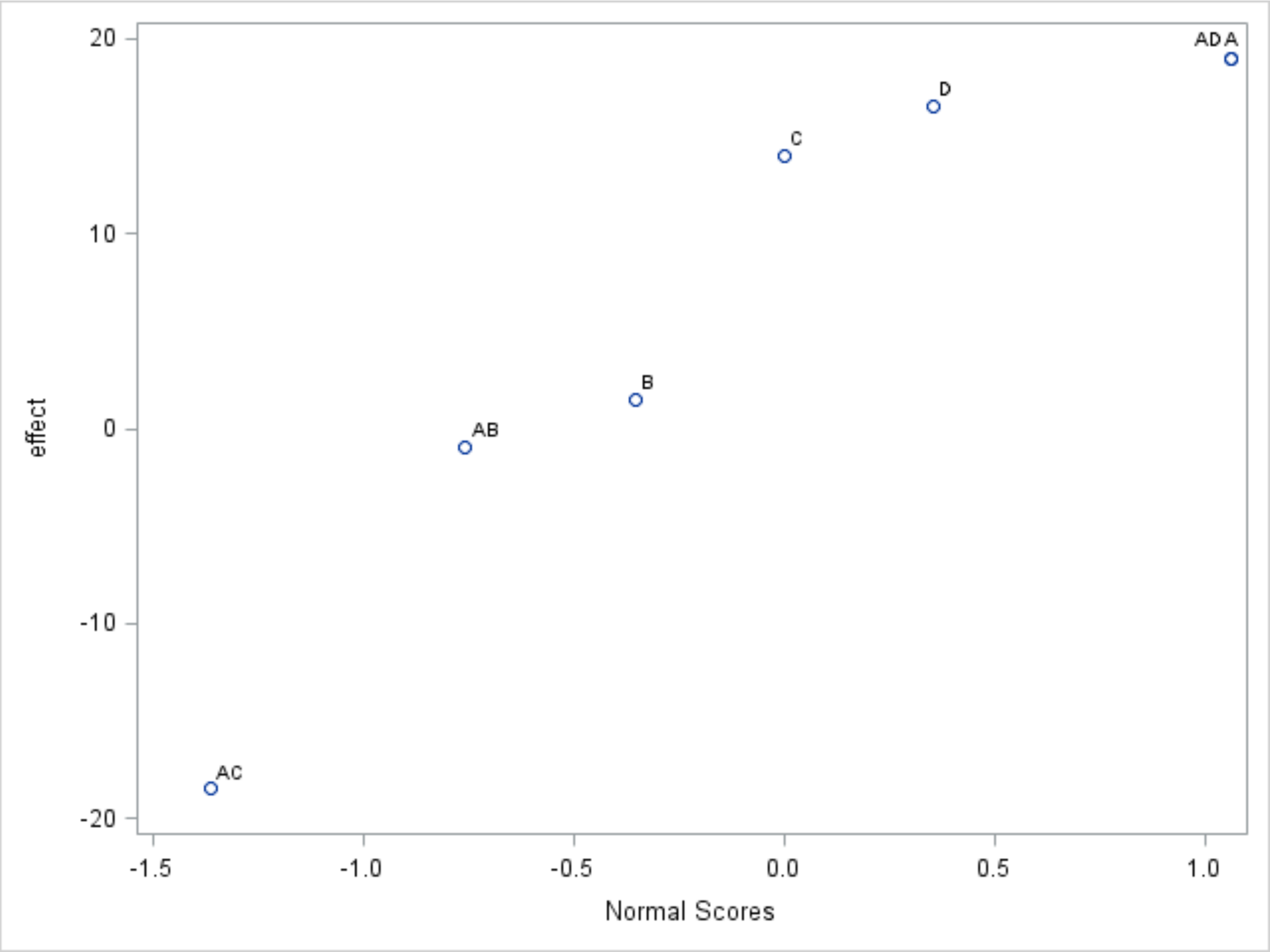
SAS Output

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	7	3071.500000	438.785714	.	.
Error	0	0.000000	.		
CoTotal	7	3071.500000			

Source	DF	Type I SS	Mean Square	F Value	Pr > F
A	1	722.0000000	722.0000000	.	.
B	1	4.5000000	4.5000000	.	.
C	1	392.0000000	392.0000000	.	.
D	1	544.5000000	544.5000000	.	.
AB	1	2.0000000	2.0000000	.	.
AC	1	684.5000000	684.5000000	.	.
AD	1	722.0000000	722.0000000		

Obs	_NAME_	COL1	effect
1	AC	-9.25	-18.5
2	AB	-0.50	-1.0
3	B	0.75	1.5
4	C	7.00	14.0
5	D	8.25	16.5
6	A	9.50	19.0
7	AD	9.50	19.0

QQ plot to identify important effects



Potentially important effects: A, C, D, AC, AD

Regression Model

Let x_1, x_3, x_4 be the variables for factor A, C and D. The model is

$$y = 70.75 + 9.50x_1 + 7.00x_3 + 8.25x_4 - 9.25x_1x_3 + 9.50x_1x_4$$

In Lecture 14, the regression model based on all the data (2^4) is

$$y = 70.06 + 10.81x_1 + 4.94x_3 + 7.31x_4 - 9.06x_1x_3 + 8.31x_1x_4$$

It appears that the model based on 2^{4-1} is as good as the original one.

Is this really true? The answer is **NO**, because the chosen effects are aliased with other effects, so we have to resolve the ambiguities between the aliased effects first.

Aliased effects and Techniques for Resolving the Ambiguities

The estimates are for the sum of aliased factorial effects.

$$\mathcal{L}_I = 70.75 \to I + ABCD$$

$$\mathcal{L}_A = 19.0 \to A + BCD$$

$$\mathcal{L}_B = 1.5 \to B + ACD$$

$$\mathcal{L}_C = 14.0 \to C + ABD$$

$$\mathcal{L}_D = 16.5 \to D + ABC$$

$$\mathcal{L}_{AB} = -1.0 \to AB + CD$$

$$\mathcal{L}_{AC} = -18.5 \to AC + BD$$

$$\mathcal{L}_{AD} = 19.0 \to AD + BC$$

Techniques for resolving the ambiguities in aliased effects

- Use the fundamental principles (Slide 1)
- Follow-up Experiment
  - add orthogonal runs, or optimal design approach, or fold-over design

Sequential Experiment

If it is necessary, the remaining 8 runs of the original 2^4 design can be conducted.

- Recall that the 8 runs we have used are defined by $I = ABCD$. The remaining 8 runs are indeed defined by the following relationship

$$D = -ABC, \text{ or } I = -ABCD$$

basic design				
A	B	C	$D = -ABC$	filtration rate
$-$	$-$	$-$	$+$	43
$+$	$-$	$-$	$-$	71
$-$	$+$	$-$	$-$	48
$+$	$+$	$-$	$+$	104
$-$	$-$	$+$	$-$	68
$+$	$-$	$+$	$+$	86
$-$	$+$	$+$	$+$	70
$+$	$+$	$+$	$-$	65

$I = -ABCD$ implies that: $A = -BCD$, $B = -ACD$, $\ldots$, $AB = -CD \ldots$

Similarly, we can derive the following estimates ($\tilde{\mathcal{L}}_{\text{effect}}$) and alias structure

$$
\begin{array}{lcl}
\tilde{\mathcal{L}}_I = 69.375 & \rightarrow & I - ABCD \\
\tilde{\mathcal{L}}_A = 24.25 & \rightarrow & A - BCD \\
\tilde{\mathcal{L}}_B = 4.75 & \rightarrow & B - ACD \\
\tilde{\mathcal{L}}_C = 5.75 & \rightarrow & C - ABD \\
\tilde{\mathcal{L}}_D = 12.75 & \rightarrow & D - ABC \\
\tilde{\mathcal{L}}_{AB} = 1.25 & \rightarrow & AB - CD \\
\tilde{\mathcal{L}}_{AC} = -17.75 & \rightarrow & AC - BD \\
\tilde{\mathcal{L}}_{AD} = 14.25 & \rightarrow & AD - BC
\end{array}
$$

Combine Sequential Experiments

Combining two experiments $\Rightarrow 2^4$ full factorial experiment

Combining the estimates from these two experiments $\Rightarrow$ estimates based on the full experiment

$$\mathcal{L}_A = 19.0 \rightarrow A + BCD$$

$$\tilde{\mathcal{L}}_A = 24.25 \rightarrow A - BCD$$

$$A = \frac{1}{2}(\mathcal{L}_A + \tilde{\mathcal{L}}_A) = 21.63$$

$$BCD = \frac{1}{2}(\mathcal{L}_A - \tilde{\mathcal{L}}_A) = -2.63$$

Other effects are summarized in the following table

i	$\frac{1}{2}(\mathcal{L}_i + \tilde{\mathcal{L}}_i)$	$\frac{1}{2}(\mathcal{L}_i - \tilde{\mathcal{L}}_i)$
A	21.63 $\rightarrow A$	-2.63 $\rightarrow BCD$
B	3.13 $\rightarrow B$	-1.63 $\rightarrow ACD$
C	9.88 $\rightarrow C$	4.13 $\rightarrow ABD$
D	14.63 $\rightarrow D$	1.88 $\rightarrow ABC$
AB	.13 $\rightarrow AB$	-1.13 $\rightarrow CD$
AC	-18.13 $\rightarrow AC$	-0.38 $\rightarrow BD$
AD	16.63 $\rightarrow AD$	2.38 $\rightarrow BC$

We know the combined experiment is not a completely randomized experiment. Is there any underlying factor we need consider? what is it?

General 2^{k-1} Design

- k factors: A, B, . . ., K
- can only afford half of all the combinations (2^{k-1})
- Basic design: a 2^{k-1} full factorial for $k-1$ factors: A, B, . . ., J.
- The setting of kth factor is determined by alasing K with the $ABC....J$, i.e., $K = ABC \cdots J$
- Defining relation: $I = ABCD....\tilde{I}JK$. Resolution=k
- 2^k factorial effects are partitioned into 2^{k-1} groups each with two aliased effects.
- only one effect from each group (the representative) should be included in ANOVA or regression model.
- Use fundamental principles, domain knowledge, follow-up experiment to de-alias.

One Quarter Fraction: 2^{k-2} Design

Parts manufactured in an injection molding process are showing excessive shrinkage. A quality improvement team has decided to use a designed experiment to study the injection molding process so that shrinkage can be reduced. The team decides to investigate six factors

A: mold temperature

B: screw speed

C: holding time

D: cycle time

E: gate size

F: holding pressure

each at two levels, with the objective of learning about main effects and interactions. They decide to use 16-run fractional factorial design.

- a full factorial has 2^6=64 runs.
- 16-run is one quarter of the full factorial
- How to construct the fraction?

Injection Molding Experiment: 2^{6-2} Design

basic design						
A	B	C	D	$E = ABC$	$F = BCD$	shrinkage
−	−	−	−	−	−	6
+	−	−	−	+	−	10
−	+	−	−	+	+	32
+	+	−	−	−	+	60
−	−	+	−	+	+	4
+	−	+	−	−	+	15
−	+	+	−	−	−	26
+	+	+	−	+	−	60
−	−	−	+	−	+	8
+	−	−	+	+	+	12
−	+	−	+	+	−	34
+	+	−	+	−	−	60
−	−	+	+	+	−	16
+	−	+	+	−	−	5
−	+	+	+	−	+	37
+	+	+	+	+	+	52

Two defining relations are used to generate the columns for E and F.

$$I = ABCE, \text{ and } I = BCDF$$

They induce another defining relation:

$$I = ABCE * BCDF = AB^2C^2DEF = ADEF$$

The complete defining relation:

$$I = ABCE = BCDF = ADEF$$

Defining contrasts subgroup: $\{I, ABCE, BCDF, ADEF\}$

Alias Structure for 2^{6-2} with $I = ABCE = BCDF = ADEF$

$I = ABCE = BCDF = ADEF$ implies

$$A = BCE = ABCDF = DEF$$

Similarly, we can derive the other groups of aliased effects.

$A = BCE = DEF = ABCDF$	$AB = CE = ACDF = BDEF$
$B = ACE = CDF = ABDEF$	$AC = BE = ABDF = CDEF$
$C = ABE = BDF = ACDEF$	$AD = EF = BCDE = ABCF$
$D = BCF = AEF = ABCDE$	$AE = BC = DF = ABCDEF$
$E = ABC = ADF = BCDEF$	$AF = DE = BCEF = ABCD$
$F = BCD = ADE = ABCEF$	$BD = CF = ACDE = ABEF$
	$BF = CD = ACEF = ABDE$
$ABD = CDE = ACF = BEF$	
$ACD = BDE = ABF = CEF$	

Wordlength pattern $W = (W_0, W_1, \ldots, W_6)$, where W_i is the number of defining words of length i (i.e., involving i factors)

$$W = (1, 0, 0, 0, 3, 0, 0)$$

Resolution is the smallest i such that $i > 0$ and $W_i > 0$. Hence it is a 2^{6-2}_{IV} design

2^{6-2} Design: an Alternative

- Basic Design: A, B, C, D
- $E = ABCD, F = ABC$, i.e., $I = ABCDE$, and $I = ABCF$
- which induces: $I = DEF$
- complete defining relation: $I = ABCDE = ABCF = DEF$
- wordlength pattern: $W = (1, 0, 0, 1, 1, 1, 0)$

Maximum possible resolution

- Alias structure (ignore effects of order 3 or higher)

$A = ..$	$AB = CF = ..$
$B = ..$	$AC = BF = ..$
$C = ..$	$AD = ..$
$D = EF = ..$	$AE = ..$
$E = DF = ..$	$AF = BC = ..$
$F = DE = ..$	$BD = ..$
	$BE = ..$
	$CD = ..$
	$CE = ..$

- an effect is said to be **clearly estimable** if it is not aliased with main effect or two-factor interactions.
- Which design is better d_1 or d_2? d_1 has six clearly estimable main effects while d_2 has three clearly estimable main effects and six clearly estimable two-factor ints.

SAS code: Injection Molding Experiment Analysis

```
goption colors=(none);
data molding;
 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; E=A*B*C; F=B*C*D;
 input y @@;  output; end; end; end; end;
 datalines;
 6 10 32 60 4 15 26 60 8 12 34 60 16 5 37 52
;
data inter;       /* Define Interaction Terms */
set molding;
AB=A*B; AC=A*C; AD=A*D; AE=A*E; AF=A*F; BD=B*D; BF=B*F; ABD=A*B*D;
ACD=A*C*D;

proc glm data=inter;    /* GLM Proc to Obtain Effects */
class A B C D E F AB AC AD AE AF BD BF ABD ACD;
model y=A B C D E F AB AC AD AE AF BD BF ABD ACD;
estimate 'A' A -1 1; estimate 'B' B -1 1; estimate 'C' C -1 1;
estimate 'D' D -1 1; estimate 'E' E -1 1; estimate 'F' F -1 1;
```

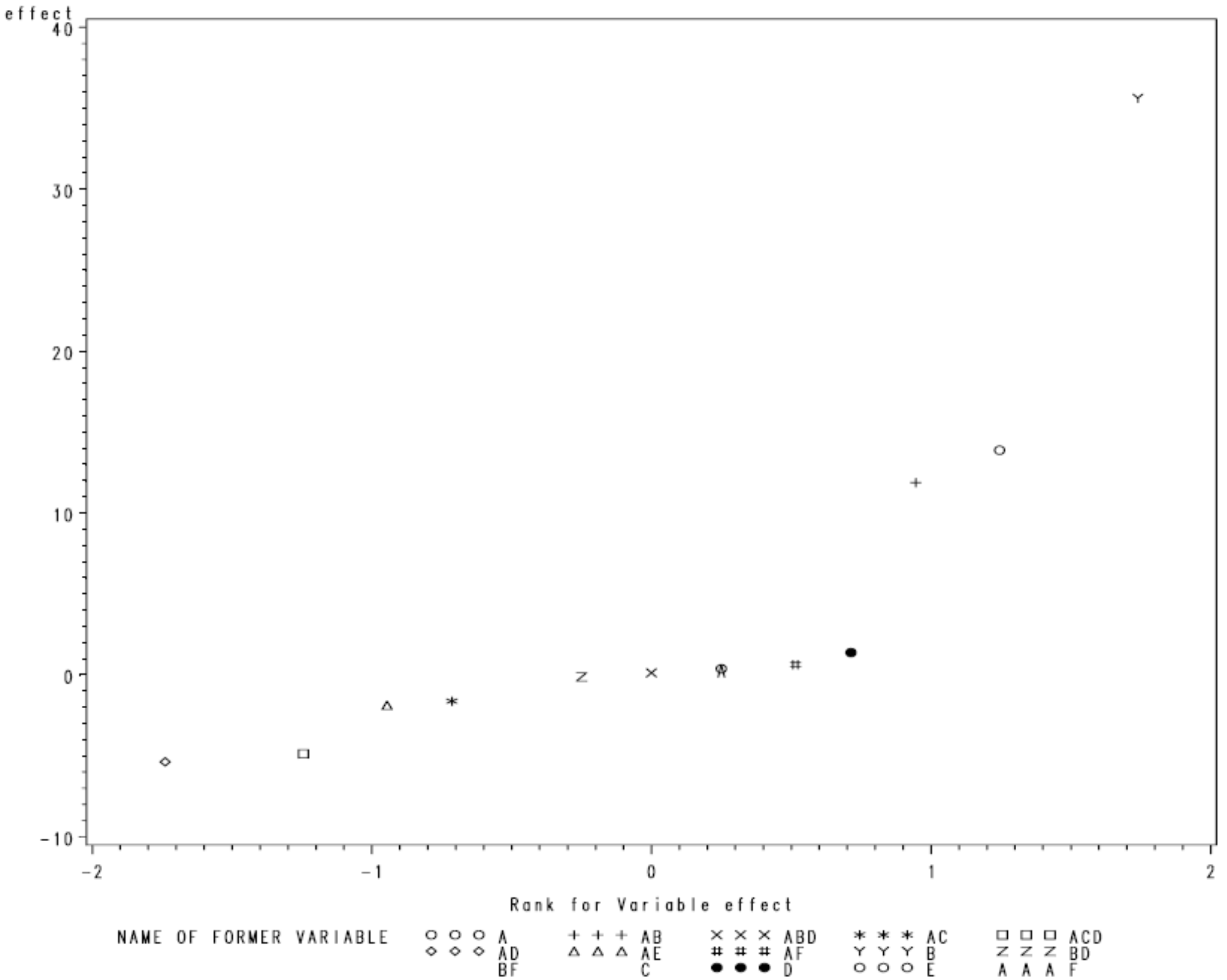
```
estimate 'AB' AB -1 1; estimate 'AC' AC -1 1; estimate 'AD' AD -1 1;
estimate 'AE' AE -1 1; estimate 'AF' AF -1 1; estimate 'BD' BD -1 1;
estimate 'BF' BF -1 1; estimate 'ABD' ABD -1 1; estimate 'ACD' ACD -1 1;
run;

proc reg outest=effects data=inter;    /* REG Proc to Obtain Effects */
model y=A B C D E F AB AC AD AE AF BD BF ABD ACD;
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;
proc rank data=effect4 normal=blom; var effect; ranks neff;
symbol1 v=circle;
proc gplot; plot effect*neff=_NAME_; run;
```

Estimates of factorial effects

Obs	_NAME_	COL1	effect	aliases
1	AD	-2.6875	-5.375	AD+EF
2	ACD	-2.4375	-4.875	
3	AE	-0.9375	-1.875	AE+BC+DF
4	AC	-0.8125	-1.625	AC+BE
5	C	-0.4375	-0.875	
6	BD	-0.0625	-0.125	BD+CF
7	BF	-0.0625	-0.125	BF+CD
8	ABD	0.0625	0.125	
9	E	0.1875	0.375	
10	F	0.1875	0.375	
11	AF	0.3125	0.625	AF+DE
12	D	0.6875	1.375	
13	AB	5.9375	11.875	AB+CE
14	A	6.9375	13.875	
15	B	17.8125	35.625	

Effects B, A, AB, AD, ACD, are large.

QQ plot to Identify Important Effects



Effects B, A, AB appear to be important; effects AD and ACD are suspicious.

De-aliasing and Model Selection

```
Model 1:
proc reg data=inter;
model y=A B AB AD ACD;
run;
--------------------
     Root MSE              1.95256    R-Square     0.9943
     Dependent Mean       27.31250    Adj R-Sq     0.9914
     Coeff Var             7.14897
                     Parameter Estimates
                     Parameter       Standard
Variable      DF     Estimate          Error    t Value  Pr > |t|
Intercept      1     27.31250        0.48814      55.95    <.0001
A              1      6.93750        0.48814      14.21    <.0001
B              1     17.81250        0.48814      36.49    <.0001
AB             1      5.93750        0.48814      12.16    <.0001
AD             1     -2.68750        0.48814      -5.51    0.0003
ACD            1     -2.43750        0.48814      -4.99    0.0005
============================================
```

```
Model 2:
proc reg data=inter;
model y=A B AB;
------------------------
```

Root MSE	4.55293	R-Square	0.9626
Dependent Mean	27.31250	Adj R-Sq	0.9533
Coeff Var	16.66976		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	27.31250	1.13823	24.00	<.0001
A	1	6.93750	1.13823	6.09	<.0001
B	1	17.81250	1.13823	15.65	<.0001
AB	1	5.93750	1.13823	5.22	0.0002

General 2^{k-p} Fractional Factorial Designs

- k factors, 2^k level combinations, but want to run a 2^{-p} fraction only.
- Select the first $k-p$ factors to form a full factorial design (basic design).
- Alias the remaining p factors with some high order interactions of the basic design.
- There are p defining relation, which induces other $2^p - p - 1$ defining relations. The complete defining relation is $I = .. = ... = ...$
- Defining contrasts subgroup: $G = \{$ defining words$\}$
- Wordlength pattern: $W = (W_i)$ W_i=the number of defining words of length i.

- Alias structure: 2^k factorial effects are partitioned into 2^{k-p} groups of effects, each of which contains 2^p effects. Effects in the same group are aliased (aliases).
- Use maximum resolution and minimum aberration to choose the optimal design.
- In analysis, only select one effect from each group to be included in the full model.
- Choose important effect to form models, pool unimportant effects into error component
- De-aliasing and model selection.

Minimum Aberration Criterion

Recall 2^{k-p} with maximum resolution should be preferred. But, it is possible that there are two designs that attain the maximum resolution. How should we further distinguish them?

For example, consider 2^{7-2} fractional factorial design

d_1: basic design: $A, B, C, D, E; \quad F = ABC, G = ABDE$

complete defining relation: $I = ABCF = ABDEG = CDEFG$

wordlength pattern: $W = (1, 0, 0, 0, 1, 2, 0, 0)$

Resolution: IV

d_2: basic design: $A, B, C, D, E; \quad F = ABC, G = ADE$

complete defining relation: $I = ABCF = ADEG = BCDEFG$

wordlength pattern: $W = (1, 0, 0, 0, 2, 0, 1, 0)$

Resolution: IV

d_1 and d_2, which is better?

Minimum Aberration Criterion

Definition: Let d_1 and d_2 be two 2^{k-p} designs, let r be the smallest **positive** integer such that $W_r(d_1) \neq W_r(d_2)$.

If $W_r(d_1) < W_r(d_2)$, then d_1 is said to have less aberration than d_2.

If there does

not exist any other design that has less aberration than d_1, then d_1 has minimum aberration.

Small Minimum Aberration Designs are used a lot in practice. They are tabulated in most design books. See Table 8-14 in Montgomery.

Last slide

- Read Sections: 8.1- 8.7