清华大学统计学辅修课程

Design and Analysis of Experiments

Lecture 8 –
2^{k-p} Fractional Factorial Design

周在莹
清华大学统计学研究中心
http://www.stat.tsinghua.edu.cn

Outline

- Fundamental Principles
- Fractional Factorial designs- one of the most important designs for screening
- 别名 Aliasing, Defining Relation and Word
  - design generator, alias structure, word length pattern
- 辨识度 Design Resolution and Aberration 低阶混杂
  - maximum resolution and minimum aberration
- Fold-Over Technique
- Plackett-Burman Designs

Fundamental Principles Regarding Factorial Effects

- Suppose there are k factors ($A,B,\ldots,J,K$) 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 阶层有序
  - 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) 稀疏性原则
  - 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

Motivation

- There may be **many** variables (often because we don't know much about the system). Need $r2^k$ runs for k factors (r= # of replicates)
- As the number of factors becomes large enough to be "interesting", the size of the designs grows very quickly, detailed later
- Emphasis is on **factor screening**; efficiently identify the factors with large effects
- May not have sources (time, money, etc) for full factorial design

Fraction Is Enough?

- Number of runs required for full factorial grows quickly, even r =1
  - If k = 7 ⇒ 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
- Almost always run as **unreplicated** factorials, but often with **center points**

Discussion: 2^5 design

- How many main effects and interaction effects respectively?

	Main	Interactions			
	Effects	2-Factor	3-Factor	4-Factor	5-Factor
#	5	10	10	5	1

- How many degrees of freedom for this design?
  - 31 degrees of freedom in a 2^5 design
- A full factorial design
  - Covers all main effects and interaction effects
  - Uses most degrees of freedom for interaction effects
- Use of a FF design instead of full factorial design is usually done for economic reasons. Since there is no free lunch , what price to pay?

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 to investigate all nine factors in the experiment
  - A fraction of 2^9 full factorial design will be used
  - <u>Confounding (aliasing) will happen because using a subset</u>
- How to choose (or construct) the fraction?

Aliasing of effects is a price one must pay for choosing a smaller design

Example 2: Filtration Rate Experiment

- Recall that there are four factors in the experiment (A, B, C and D), each of 2 levels
- 2^4 full factorial design consists of all the 16 level combinations of the four factors
- Suppose the available resource is enough for conducting 8 runs
- We need to choose half of them
- The chosen half is called 2^{4-1} fractional factorial design
- Question: Which half we should select (construct)?

factor			
A	B	C	D
−	−	−	−
+	−	−	−
−	+	−	−
+	+	−	−
−	−	+	−
+	−	+	−
−	+	+	−
+	+	+	−
−	−	−	+
+	−	−	+
−	+	−	+
+	+	−	+
−	−	+	+
+	−	+	+
−	+	+	+
+	+	+	+

Effect Aliasing and Defining Relation

- 2^{4-1} Fractional Factorial Design
  - the number of factors: $k = 4$
  - the fraction index: $p = 1$
  - the number of runs (level combinations):

$$N = 2^4/2^1 = 8$$

- 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

Defining Relation & Defining Word

- $D = ABC$, 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)

	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

Aliasing in 2^{4-1} Design

- For four factors A, B, C and D, there are $2^4 - 1$ effects:
  - $A, B, C, D, AB, AC, AD, BC, BD, CD, ABC, ABD, ACD, BCD, ABCD$
- Contrasts for main effects by converting $-$ to -1 and $+$ to 1; contrasts for other effects obtained by multiplication

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

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

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

  $\Rightarrow$ *A* and *BCD* are aliases/aliased/not distinguishable. The contrast is for *A*+*BCD*
- *AB, CD* …There are other 5 pairs are aliases or aliased… They are caused by the defining relation

$$I = ABCD;$$

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

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

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

- Alias Structure:
  - $I = ABCD$
  - $A = A * I = A * ABCD = BCD$
  - $B = \ldots = ACD$
  - $C = \ldots = ABD$
  - $D = \ldots = ABC$
  - $AB = AB * I = AB * ABCD = CD$
  - $AC = \ldots = BD$
  - $AD = \ldots = BC$
- All 16 factorial effects for A, B, C and D are partitioned into 8 groups each with 2 aliased effects

Clear Effects

- Definition: A main effect or two-factor interaction is called clear if it is not aliased with any other main effects or two-factor interactions and strongly clear if it is not aliased with any other main effects, two-factor interactions or three-factor interactions
- A clear effect is estimable under the assumption of negligible 3-factor and higher interactions and a strongly clear effect is estimable under the weaker assumption of negligible 4-factor and higher interactions
- Question: In the 2^{4-1} design with $I = ABCD$, which effects are clear and strongly clear?
- Ans: *B, C, D, E* are clear, none is strongly clear
- We usually care about the clear effects

Another 2^{4-1} Fractional Factorial Design

- The defining relation $I = ABD$ generates a different 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 fractional 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_{IV}^{4-1}
- d_2: $I = ABC$ is a resolution III design denoted by 2_{III}^{4-1}
- 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 or 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
- <u>Maximum Resolution Criterion</u>
  - 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 L_{effect} denote the estimate of effect (based on the corresponding contrast). Because of aliasing,

  $L_I \to I + ABCD$, $L_A \to A + BCD$,

  $L_B \to B + ACD$, $L_C \to C + ABD$,

  $L_D \to D + ABC$, $L_{AB} \to AB + CD$,

  $L_{AC} \to AC + BD$, $L_{AD} \to AD + BC$

R Output


```
> model <- lm(rate ~ A + B + C + D + I(A*B) +I(A*C) + I(A*D), plant1)
> summary(model)
> anova(model)
```

We call the model saturated if the design has $k = N - 1$ variables

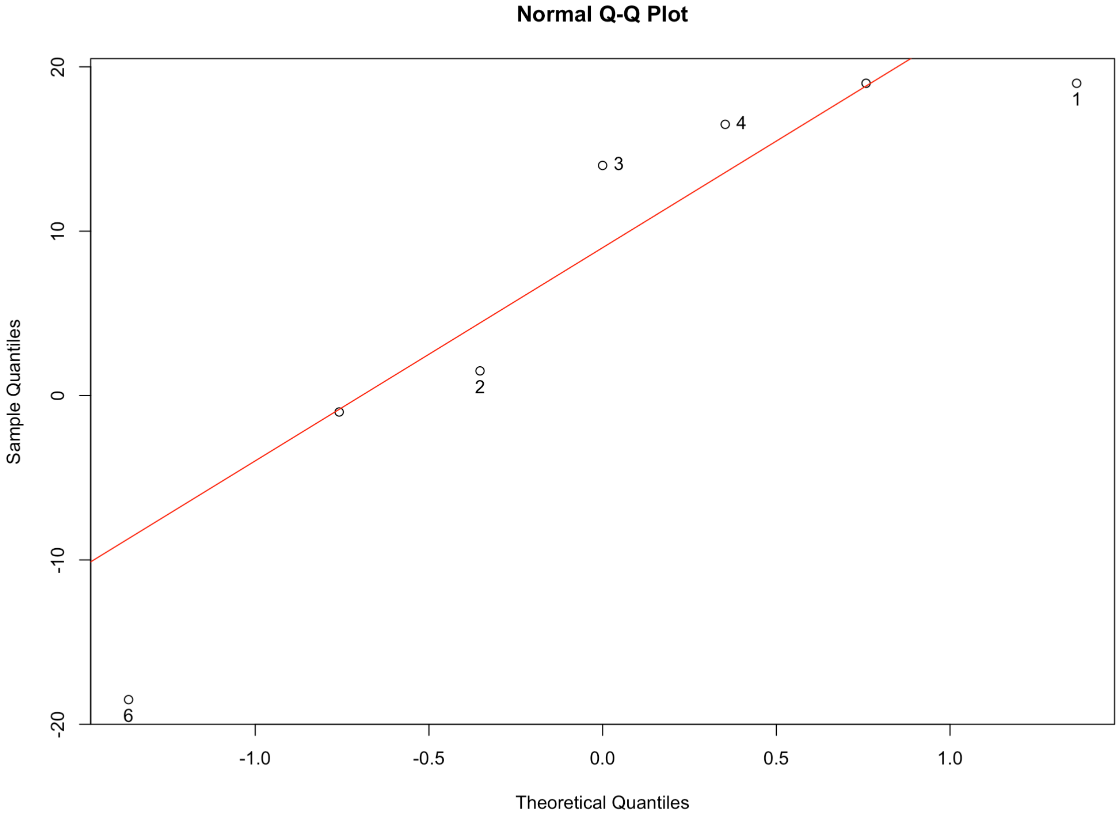
```
Analysis of Variance Table

Response: rate
          Df Sum Sq Mean Sq  F value  Pr(>F)
A          1  722.0   722.0
B          1    4.5     4.5
C          1  392.0   392.0
D          1  544.5   544.5
I(A * B)   1    2.0     2.0
I(A * C)   1  684.5   684.5
I(A * D)   1  722.0   722.0
Residuals  0    0.0
Warning message:
In anova.lm(model) :
  ANOVA F-tests on an essentially perfect fit are unreliable
```

```
Residuals:
ALL 8 residuals are 0: no residual degrees of freedom!

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  70.75      NA        NA      NA
A             9.50      NA        NA      NA
B             0.75      NA        NA      NA
C             7.00      NA        NA      NA
D             8.25      NA        NA      NA
I(A * B)     -0.50      NA        NA      NA
I(A * C)     -9.25      NA        NA      NA
I(A * D)      9.50      NA        NA      NA

Residual standard error: NaN on 0 degrees of freedom
Multiple R-squared: 1,  Adjusted R-squared: NaN
F-statistic:   NaN on 7 and 0 DF,  p-value: NA
```

QQ Plot to Identify Important Effects

Coefficients:

	Estimate	Df	Sum Sq
(Intercept)	70.75		
A	9.50	1	722.0
B	0.75	1	4.5
C	7.00	1	392.0
D	8.25	1	544.5
I(A * B)	-0.50	1	2.0
I(A * C)	-9.25	1	684.5
I(A * D)	9.50	1	722.0

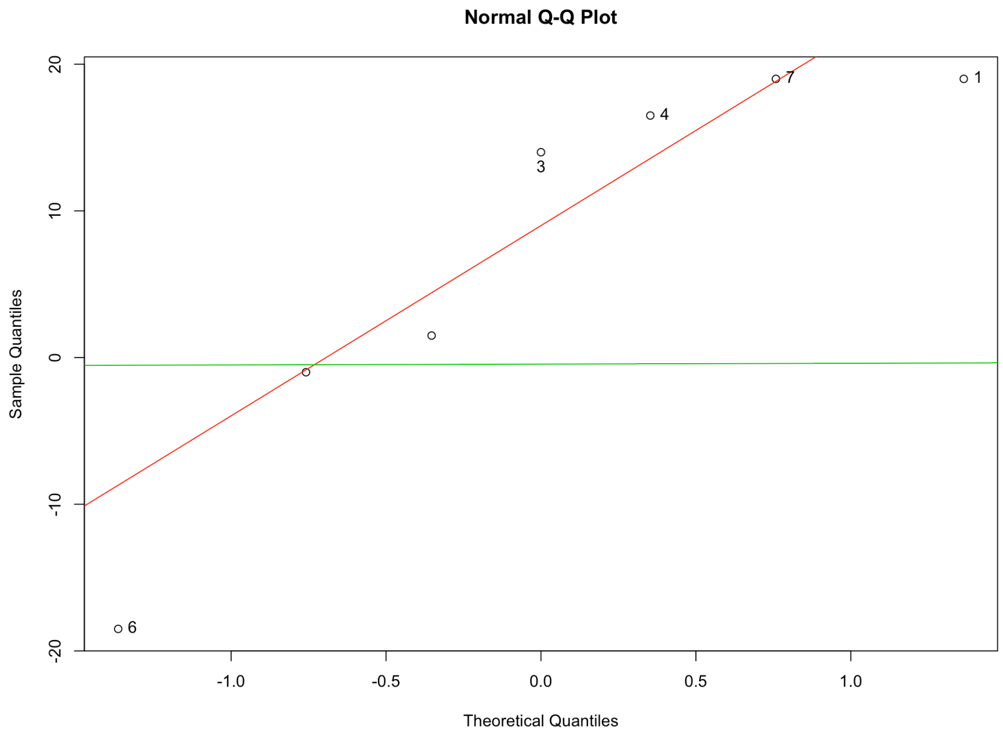


- Potentially important effects: *A, C, D, AC* and *AD*

Note: Always Check the Variances

QQ plot to Identify Important Effects

```
> effects <- (coef(model) *2)[-1] #the effect estimates
> res <- qqnorm(effects)
> qqline(effects, col = 'red')
> identify(res$x, res$y, plot = TRUE)
##realize the problem
> plot(res)
> abline(lm(res$x ~ res$y), col = 'green')
> identify(res$x, res$y, plot = TRUE)
```



Confirmation Experiment

- > summary(lm(rate ~ A + C + D + I(A*C) + I(A*D), plant1))
- Use x_1, x_3, x_4 for *A, C, D,* the regression model is
$$\hat{y} = 70.75 + 9.50x_1 + 7.00x_3 + 8.25x_4 - 9.25x_1x_3 + 9.50x_1x_4$$
- Use the model to predict the response at a test combination of interest in the design space – not one of the points in the current design
- Run this test combination – then compare predicted and observed
- For example, consider the point +, +, –, +. The predicted response is

$$\hat{y} = 70.75 + 9.50(1) + 7.00(-1) + 8.25(1) - 9.25(-1) + 9.50(1) = 100.25$$

actual response is 104

Regression Models

- Use x_1, x_3, x_4 for *A, C, D,* the regression model is
  $$\hat{y} = 70.75 + 9.50x_1 + 7.00x_3 + 8.25x_4 - 9.25x_1x_3 + 9.50x_1x_4$$
- Compared with the regression model based on all the data (2^4 design in Lec07)
  $$\hat{y} = 70.06 + 10.81x_1 + 4.94x_3 + 7.31x_4 - 9.06x_1x_3 + 8.31x_1x_4$$
- It appears that the model based on 2^{4-1} is as good as the original one
- Is this really true?
- NO, because the chosen effects are aliased with other effects, so we have to resolve the ambiguities between the aliased effects first

```
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  70.7500     0.6374  111.00 8.11e-05 ***
A             9.5000     0.6374   14.90  0.00447 **
C             7.0000     0.6374   10.98  0.00819 **
D             8.2500     0.6374   12.94  0.00592 **
I(A * C)     -9.2500     0.6374  -14.51  0.00471 **
I(A * D)      9.5000     0.6374   14.90  0.00447 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.803 on 2 degrees of freedom
Multiple R-squared:  0.9979,	Adjusted R-squared:
0.9926
F-statistic: 188.6 on 5 and 2 DF,  p-value: 0.005282
```

Aliased Effects and Techniques for Resolving the Ambiguities

- The estimates are for the sum of aliased factorial effects

$$L_I = 70.75 \rightarrow I + ABCD,$$
$$L_A = 19.0 \rightarrow A + BCD,$$
$$L_B = 1.5 \rightarrow B + ACD,$$
$$L_C = 14.0 \rightarrow C + ABD,$$
$$L_D = 16.5 \rightarrow D + ABC,$$
$$L_{AB} = -1.0 \rightarrow AB + CD,$$
$$L_{AC} = -18.5 \rightarrow AC + BD,$$
$$L_{AD} = 19.0 \rightarrow AD + BC$$

Coefficients:

	Estimate
(Intercept)	70.75
A	9.50
B	0.75
C	7.00
D	8.25
I(A * B)	-0.50
I(A * C)	-9.25
I(A * D)	9.50

Coefficients:

	Estimate	Std. Error	t value	Pr(>|t|)	
(Intercept)	70.7500	0.6374	111.00	8.11e-05	***
A	9.5000	0.6374	14.90	0.00447	**
C	7.0000	0.6374	10.98	0.00819	**
D	8.2500	0.6374	12.94	0.00592	**
I(A * C)	-9.2500	0.6374	-14.51	0.00471	**
I(A * D)	9.5000	0.6374	14.90	0.00447	**

- Techniques for resolving the ambiguities in aliased effects
  - Use the fundamental principles
  - 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 defined by $I = ABCD$. The remaining 8 runs are indeed defined by the following relationship

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

which implies that:

$$A = -BCD, B = -ACD, \ldots, AB = -CD\ldots$$

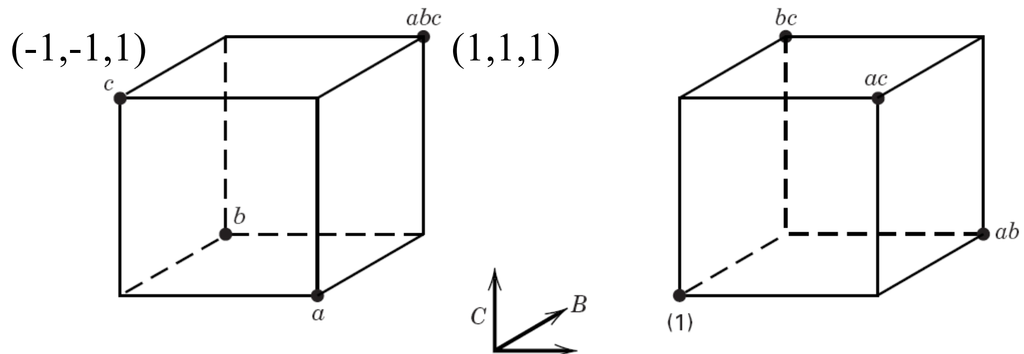
basic design				
A	B	C	$D = -ABC$	filtration rate
−	−	−	+	43
+	−	−	−	71
−	+	−	−	48
+	+	−	+	104
−	−	+	−	68
+	−	+	+	86
−	+	+	+	70
+	+	+	−	65

Note on $I = \pm ABCD$

- Both designs belong to the same **family,** defined by $I = \pm ABCD$
- One-half fraction, with $I = +ABCD$, is usually called the principal fraction
- The alternate, or complementary, one-half fraction is $I = -ABCD$

- Suppose that after running the principal fraction, the alternate fraction was also run. The two groups of runs can be combined to form a full factorial – an example of **sequential** experimentation
- Another example->

The two one-half fractions of the 2^3 design



(a) The principal fraction, $I = +ABC$

(b) The alternate fraction, $I = -ABC$

The Alias Structure

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

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

Combine Sequential Experiments

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

- Combining two experiments ⇒ the 2^4 full factorial experiment
- Combining the estimates from these two experiments ⇒ the estimates based on the full experiment
- $L_A = 19.0 \rightarrow A + BCD,\ \tilde{L}_A = 24.25 \rightarrow A - BCD$

$$\Rightarrow A = \frac{1}{2}\left(L_A + \tilde{L}_A\right) = 21.63$$

$$ABC = \frac{1}{2}\left(L_A - \tilde{L}_A\right) = -2.63$$

- Other effects are summarized in the 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$\longrightarrow A$	-2.63 $\longrightarrow BCD$
B	3.13$\longrightarrow B$	-1.63 $\longrightarrow ACD$
C	9.88$\longrightarrow C$	4.13 $\longrightarrow ABD$
D	14.63$\longrightarrow D$	1.88 $\longrightarrow ABC$
AB	.13$\longrightarrow AB$	-1.13 $\longrightarrow CD$
AC	-18.13$\longrightarrow AC$	-0.38 $\longrightarrow BD$
AD	16.63$\longrightarrow AD$	2.38 $\longrightarrow BC$

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: $A, B, \ldots, K$
- 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 aliasing K with the $ABC\ldots J$, i.e.,

$$K = ABC\ldots J$$

- Defining relation: $I = ABCD\ldots \tilde{I}JK$. 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

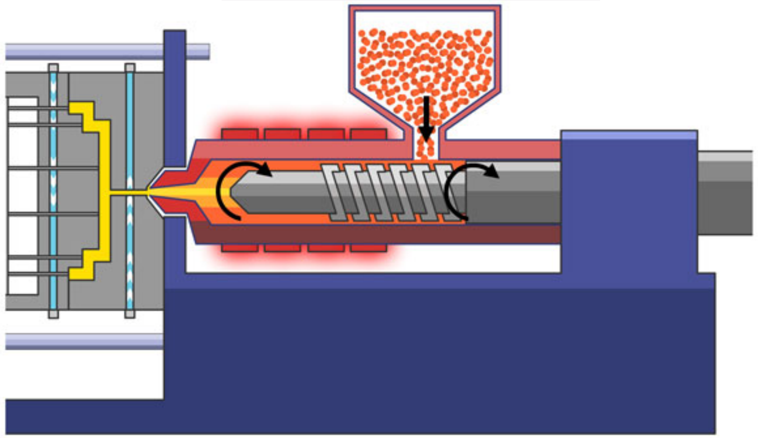
Example: Injection Molding Experiment

- 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 <u>one quarter of the full factorial</u>
- How to construct the fraction?



One Quarter Fraction: 2^{k-2} Design, d_1

- Injection Molding Experiment is a 2^{6-2} design
- Two defining relations are used to generate the columns for E and F

$$I = ABCE;\ I = BCDF$$

- They induce another defining relation:

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

- The complete defining relation:

$$I = ABCE = BCDF = ADEF$$

- Defining contrasts subgroup:

$$\{I,\ ABCE,\ BCDF,\ ADEF\}$$

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

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

- $I = ABCE = BCDF = ADEF$ implies

$$A = BCE = ABCDF = ADEF$$

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

$$\begin{aligned}
&A = BCE = DEF = ABCDF, && AB = CE = ACDF = BDEF \\
&B = ACE = CDF = ABDEF, && AC = BE = ABDF = CDEF \\
&C = ABE = BDF = ACDEF, && AD = EF = BCDE = ABCF \\
&D = BCF = AEF = ABCDE, && AE = BC = DF = ABCDEF \\
&E = ABC = ADF = BCDEF, && AF = DE = BCEF = ABCD \\
&F = BCD = ADE = ABCEF, && BD = CF = ACDE = ABEF \\
& && BF = CD = ACEF = ABDE
\end{aligned}$$

$$ABD = CDE = ACF = BEF$$

$$ACD = BDE = ABF = CEF$$

- This is a 2_{IV}^{6-2} design

Question: In this design, which effects are clear and strongly clear?

Word Length Pattern & Defining Contrast Subgroup

- For a general 2^{k-p} design, it has 2^p-1 words
- Define A_i = number of defining words of length i(i.e., involving i factors). $W = (A_3, A_4, \ldots, A_k)$ is called the word length pattern
- It is required that $A_2 = 0$ (Why?)
- Again, resolution is the shortest word length among the 2^p-1 words
- Recall that the complete defining relation:

$$I = ABCE = BCDF = ADEF$$

- Defining contrasts subgroup: The group formed by these defining words

$$\{I, ABCE, BCDF, ADEF\}$$

Note: Various Definitions on 'Word Length Pattern'

- In Montgomery, just write down the length of each word
- Some denotes $W = (A_0, A_1, \ldots, A_6)$, where A_i is the number of defining words of length i , where resolution is the smallest i such that $i > 0$ and $A_i > 0$
- For example, the 2^{6-2} design with $I = ABCE = BCDF = ADEF$
  - $W = (0, 3, 0, 0)$-here
  - $W = (4, 4, 4)$-Montgomery
  - $W = (1, 0, 0, 0, 3, 0, 0)$-else
- The resolution agrees: The design has resolution IV, it is a 2_{IV}^{6-2} design

Note: Projection Thinking

- For the 2^{6-2} Design d_1, the complete defining relation:

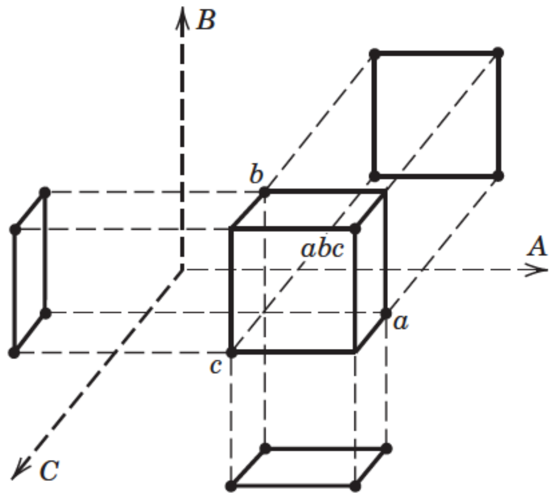
$$I = ABCE = BCDF = ADEF$$

- **Projection** of the design into subsets of the original six variables
- Any subset of the original six variables that is not a word in the complete defining relation will result in a full factorial design
  - Consider $ABCD$ (full factorial)
- Any subset of the original six variables that IS a word in the complete defining relation will result in a replicated factorial design
  - Consider $ABCE$ (replicated half fraction)

basic design					
A	B	C	D	$E = ABC$	$F = BCD$
−	−	−	−	−	−
+	−	−	−	+	−
−	+	−	−	+	+
+	+	−	−	−	+
−	−	+	−	+	+
+	−	+	−	−	+
−	+	+	−	−	−
+	+	+	−	+	−
−	−	−	+	−	+
+	−	−	+	+	+
−	+	−	+	+	−
+	+	−	+	−	−
−	−	+	+	+	−
+	−	+	+	−	−
−	+	+	+	−	+
+	+	+	+	+	+

A Projective Rationale for Resolution

- For a resolution R design, its projection onto any R-1 factors is a full factorial in the R-1 factors. This would allow effects of all orders among the R-1 factors to be estimable
- Caveat: it makes the assumption that other factors are inert
- This property can be exploited in data analysis:
- After analyzing the main effects, if only $R-1$ of them are significant, then all the interactions among the $R-1$ factors can also be estimated because the collapsed design on the $R-1$ factors is a full factorial



Projection of a 2^{3-1}_{III} design into three 2^2 designs

2^{6-2} Design: an Alternative, d_2

- Basic Design: A, B, C, D
- Generators $E = ABCD, F = ABC$, i.e.,

  $$I = ABCDE, \quad I = ABCF$$

  which induces: $I = DEF$
- Complete defining relation: $I = ABCDE = ABCF = DEF$
- Word length pattern: $W = (1, 1, 1, 0)$
- Alias structure (ignore effects of order 3 or higher) ->
- Recall that 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 interactions

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

Injection Molding Experiment Analysis

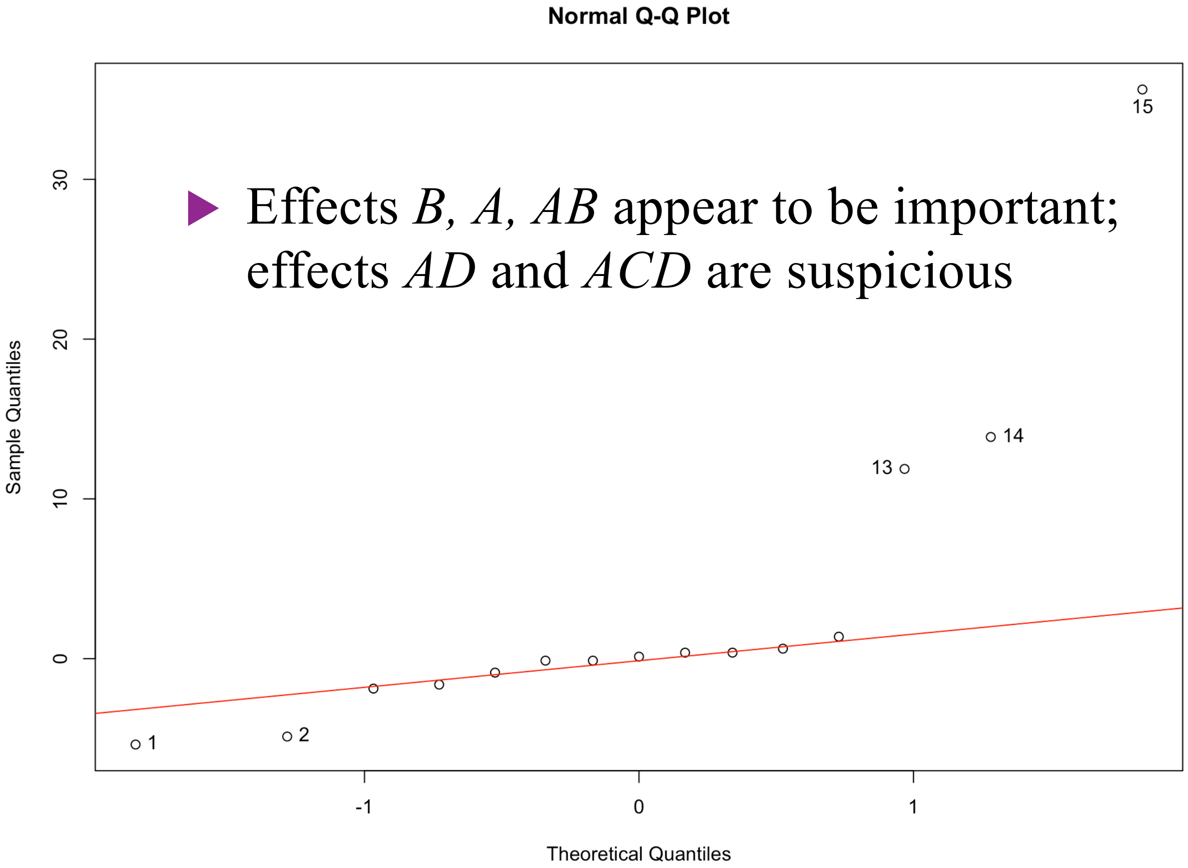
- Estimates of factorial effects
- Effcts *B, A, AB, AD, ACD*, are large

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	

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	27.3125	NA	NA	NA
A	6.9375	NA	NA	NA
B	17.8125	NA	NA	NA
C	-0.4375	NA	NA	NA
D	0.6875	NA	NA	NA
E	0.1875	NA	NA	NA
F	0.1875	NA	NA	NA
A:B	5.9375	NA	NA	NA
A:C	-0.8125	NA	NA	NA
B:C	-0.9375	NA	NA	NA
A:D	-2.6875	NA	NA	NA
B:D	-0.0625	NA	NA	NA
C:D	-0.0625	NA	NA	NA
A:E	NA	NA	NA	NA
B:E	NA	NA	NA	NA
C:E	NA	NA	NA	NA
D:E	0.3125	NA	NA	NA
A:F	NA	NA	NA	NA
…				
E:F	NA	NA	NA	NA
A:B:C	NA	NA	NA	NA
A:B:D	0.0625	NA	NA	NA
A:C:D	-2.4375	NA	NA	NA
B:C:D	NA	NA	NA	NA
……….				
A:B:C:D:E:F	NA	NA	NA	NA

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	27.3125	NA	NA	NA
A	6.9375	NA	NA	NA
B	17.8125	NA	NA	NA
C	-0.4375	NA	NA	NA
D	0.6875	NA	NA	NA
E	0.3125	NA	NA	NA
F	0.1875	NA	NA	NA
A:B	5.9375	NA	NA	NA
A:C	-0.8125	NA	NA	NA
B:C	-0.9375	NA	NA	NA
A:D	-2.6875	NA	NA	NA
B:D	-0.0625	NA	NA	NA
C:D	-0.0625	NA	NA	NA
A:E	0.1875	NA	NA	NA
B:E	-2.4375	NA	NA	NA
C:E	0.0625	NA	NA	NA
D:E	NA	NA	NA	NA
…				
A:B:C:D:E:F	NA	NA	NA	NA

QQ Plot to Identify Important Effects



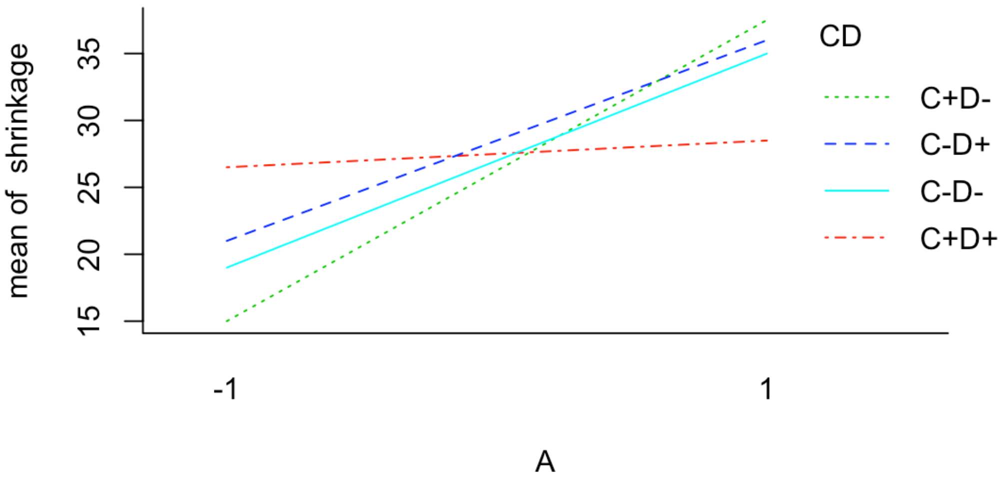
- Effects *B, A, AB* appear to be important; effects *AD* and *ACD* are suspicious

A:D	A:C:D	B:C	A:C
-5.375	-4.875	-1.875	-1.625
C	B:D	C:D	A:B:D
-0.875	-0.125	-0.125	0.125
F	E	D:E	D
0.375	0.375	0.625	1.375
A:B	A	B	
11.875	13.875	35.625	

De-aliasing and Model Selection


```
> m1 <- lm(shrinkage ~ A + B + I(A*B) + I(A*D) + I(A*C*D), mold1)
> m2 <- lm(shrinkage ~ A + B + I(A*B), mold1)
> anova(m2, m1)
```

- Three-Factor Interaction



Analysis of Variance Table

Model 1: shrinkage ~ A + B + I(A * B)
Model 2: shrinkage ~ A + B + I(A * B) + I(A * D) + I(A * C * D)

	Res.Df	RSS	Df	Sum of Sq	F	Pr(>F)	
1	12	248.750					
2	10	38.125	2	210.62	27.623	8.457e-05	***

Choosing a Design

- 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?

A design cannot be judged by its resolution alone

- For example, consider 2^{7-2} fractional factorial design
  - d_1: basic design: A, B, C, D, E; $F = ABC$, $G = ABDE$

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

    word length pattern: $W = (0, 1, 2, 0, 0)$

    Resolution: IV
  - d_2: basic design: A, B, C, D, E; $F = ABC$, $G = ADE$

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

    word length pattern: $W = (0, 2, 0, 1, 0)$

    Resolution: IV
- d_1 and d_2, which is better?

Intuitively one would argue that d_1 is better because

$$A_4(d_1) = 1 < A_4(d_2) = 2$$

(Why? Effect hierarchy principle.)

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 $Ar(d_1) \neq Ar(d_2)$
- If $A_r(d_1) < A_r(d_2)$, 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
- $W(d_1) = (0, 1, 2, 0, 0)$; $W(d_2) = (0, 2, 0, 1, 0)$
- Minimizing aberration in a design of resolution R ensures that the design has the minimum number of main effects aliased with interactions of order $R - 1$, the minimum number of two-factor interactions aliased with interactions of order $R - 2$, and so forth
- For given k and p, a minimum aberration design always exists
- Small Minimum Aberration Designs are used a lot in practice. They are tabulated in most design books. See Table 8-14 in Montgomery. For the most comprehensive table, consult Wu & Hamada

16-Run 2^{k-p} FFD $(k-p=4)$

- k is the number of factors and F&R is the fraction and resolution

k	F&R	Design Generators	Clear Effects
5	2_V^{5-1}	$5=1234$	all five main effects, all 10 2fi's
6	2_{IV}^{6-2}	$5=123, 6=124$	all six main effects
6^*	2_{III}^{6-2}	$5=12, 6=134$	3, 4, 6, 23, 24, 26, 35, 45, 56
7	2_{IV}^{7-3}	$5=123, 6=124, 7=134$	all seven main effects
8	2_{IV}^{8-4}	$5=123, 6=124, 7=134, 8=234$	all eight main effects
9	2_{III}^{9-5}	$5=123, 6=124, 7=134, 8=234, 9=1234$	none
10	2_{III}^{10-6}	$5=123, 6=124, 7=134, 8=234, 9=1234, t_0=34$	none
11	2_{III}^{11-7}	$5=123, 6=124, 7=134, 8=234, 9=1234, t_0=34, t_1=24$	none
12	2_{III}^{12-8}	$5=123,\ 6=124,\ 7=134,\ 8=234,\ 9=1234,\ t_0=34,\ t_1=24, t_2=14$	none
13	2_{III}^{13-9}	$5=123,\ 6=124,\ 7=134,\ 8=234,\ 9=1234,\ t_0=34,\ t_1=24, t_2=14, t_3=23$	none
14	2_{III}^{14-10}	$5=123,\ 6=124,\ 7=134,\ 8=234,\ 9=1234,\ t_0=34,\ t_1=24, t_2=14, t_3=23, t_4=13$	none
15	2_{III}^{15-11}	$5=123,\ 6=124,\ 7=134,\ 8=234,\ 9=1234,\ t_0=34,\ t_1=24, t_2=14, t_3=23, t_4=13, t_5=12$	none

32-Run 2^{k-p} FFD ($k-p=5$, $6\leq k\leq 11$)

- k is the number of factors and F&R is the fraction and resolution

k	F&R	Design Generators	Clear Effects
6	2_{VI}^{6-1}	$6=12345$	all six main effects, all 15 2fi's
7	2_{IV}^{7-2}	$6=123, 7=1245$	all seven main effects, 14, 15, 17, 24, 25, 27, 34, 35, 37, 45, 46, 47, 56, 57, 67
8	2_{IV}^{8-3}	$6=123, 7=124, 8=1345$	all eight main effects, 15, 18, 25, 28, 35, 38, 45, 48, 56, 57, 58, 68, 78
9	2_{IV}^{9-4}	$6=123, 7=124, 8=125, 9=1345$	all nine main effects, 19, 29, 39, 49, 59, 69, 79, 89
9	2_{IV}^{9-4}	$6=123, 7=124, 8=134, 9=2345$	all nine main effects, 15, 19, 25, 29, 35, 39, 45, 49, 56, 57, 58, 59, 69, 79, 89
10	2_{IV}^{10-5}	$6=123, 7=124, 8=125, 9=1345, t_0=2345$	all 10 main effects
10	2_{III}^{10-5}	$6=12, 7=134, 8=135, 9=145, t_0=345$	$3, 4, 5, 7, 8, 9, t_0, 23, 24, 25, 27, 28, 29, 2t_0, 36, 46, 56, 67, 68, 69, 6t_0$
11	2_{IV}^{11-6}	$6=123, 7=124, 8=134, 9=125, t_0=135, t_1=145$	all 11 main effects
11	2_{III}^{11-6}	$6=12, 7=13, 8=234, 9=235, t_0=245, t_1=1345$	$4, 5, 8, 9, t_0, t_1, 14, 15, 18, 19, 1t_0, 1t_1$

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 relations, which induce other $2^p - p - 1$ defining relations. The complete defining relation is $I = \ldots = \ldots = \ldots$
- Defining contrasts subgroup: $G = \{\text{defining words}\}$
- Word length 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

Choice of Fractions and Avoidance of Specific Combinations

- A 2^{k-p} design has 2^p choices. In general, use randomization to choose one of them. For example, the 2^{6-3} design has 8 choices

$$4 = \pm 12,\ 5 = \pm 13,\ 6 = \pm 23$$

  Randomly choose the signs

- If specific combinations (e.g., (+++) for high pressure, high temperature, high concentration) are deemed undesirable or even disastrous, they can be avoided by choosing a fraction that does not contain them. Example on p.237 of WH

Main Effects Model

- Only main effects are considered
- All interaction effects are ignored

A 2^{7-4} design for a main effects model with 7 factors

	Basic Design						
Run	A	B	C	$D = AB$	$E = AC$	$F = BC$	$G = ABC$
1	−	−	−	+	+	+	−
2	+	−	−	−	−	+	+
3	−	+	−	−	+	−	+
4	+	+	−	+	−	−	−
5	−	−	+	+	−	−	+
6	+	−	+	−	+	−	−
7	−	+	+	−	−	+	−
8	+	+	+	+	+	+	+

Example: Leaf Spring Experiment

- y: free height of spring, target is 8.0 inches
  - Goal: get y as close to 8.0 as possible (nominal-the-best problem)
- Five factors at two levels, use a 16-run design with three replicates for each run. It is a 2^{5-1} design, 1/2 fraction of the 2^5 design

Factor		Level −	Level +
B.	high heat temperature (°F)	1840	1880
C.	heating time (seconds)	23	25
D.	transfer time (seconds)	10	12
E.	hold down time (seconds)	2	3
Q.	quench oil temperature (°F)	130-150	150-170

Leaf Spring Experiment: Design Matrix and Data

- $E=BCD$
- Two-Step Procedure for Nominal-the-Best Problem
  - (i) Select levels of some factors to minimize Var(y)
  - (ii) Select the level of a factor not in (i) to move E(y) closer to the target

Factor					Free Height					
B	C	D	E	Q				$\bar{y}_i$	s_i^2	$\ln s_i^2$
−	+	+	−	−	7.78	7.78	7.81	7.7900	0.0003	-8.1117
+	+	+	+	−	8.15	8.18	7.88	8.0700	0.0273	-3.6009
−	−	+	+	−	7.50	7.56	7.50	7.5200	0.0012	-6.7254
+	−	+	−	−	7.59	7.56	7.75	7.6333	0.0104	-4.5627
−	+	−	+	−	7.94	8.00	7.88	7.9400	0.0036	-5.6268
+	+	−	−	−	7.69	8.09	8.06	7.9467	0.0496	-3.0031
−	−	−	−	−	7.56	7.62	7.44	7.5400	0.0084	-4.7795
+	−	−	+	−	7.56	7.81	7.69	7.6867	0.0156	-4.1583
−	+	+	−	+	7.50	7.25	7.12	7.2900	0.0373	-3.2888
+	+	+	+	+	7.88	7.88	7.44	7.7333	0.0645	-2.7406
−	−	+	+	+	7.50	7.56	7.50	7.5200	0.0012	-6.7254
+	−	+	−	+	7.63	7.75	7.56	7.6467	0.0092	-4.6849
−	+	−	+	+	7.32	7.44	7.44	7.4000	0.0048	-5.3391
+	+	−	−	+	7.56	7.69	7.62	7.6233	0.0042	-5.4648
−	−	−	−	+	7.18	7.18	7.25	7.2033	0.0016	-6.4171
+	−	−	+	+	7.81	7.50	7.59	7.6333	0.0254	-3.6717

Leaf Spring Experiment: Factorial Effects

- Analysis for Location Effects
- Same strategy as in full factorial experiments except for the interpretation and handling of aliased effects

- $E=BCD \Rightarrow I=BCDE$

  B=CDE, C=BDE, D=BCE, E= BCD,

  BC= DE, BD = CE, BE= CD,

  - Q=BCDEQ ⇒

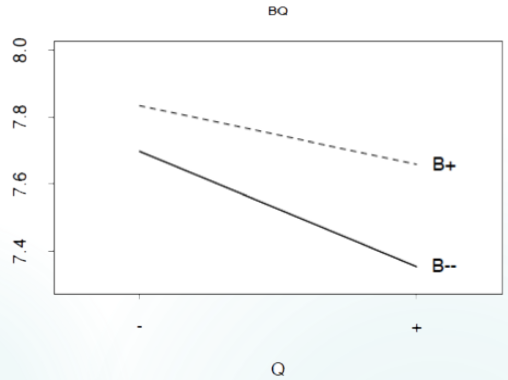
  BQ=CDEQ, CQ=BDEQ, DQ=BCEQ, EQ=BCDQ,

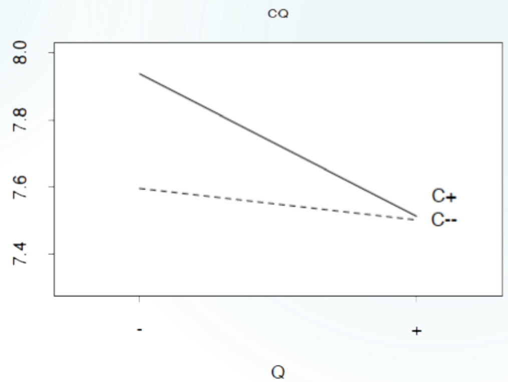
  BCQ = DEQ, BDQ = CEQ, BEQ = CDQ

Effect	$\bar{y}$	$\ln s^2$
B	0.221	1.891
C	0.176	0.569
D	0.029	-0.247
E	0.104	0.216
Q	-0.260	0.280
BQ	0.085	-0.589
CQ	-0.165	0.598
DQ	0.054	1.111
EQ	0.027	0.129
BC	0.017	-0.002
BD	0.020	0.425
CD	-0.035	0.670
BCQ	0.010	-1.089
BDQ	-0.040	-0.432
BEQ	-0.047	0.854

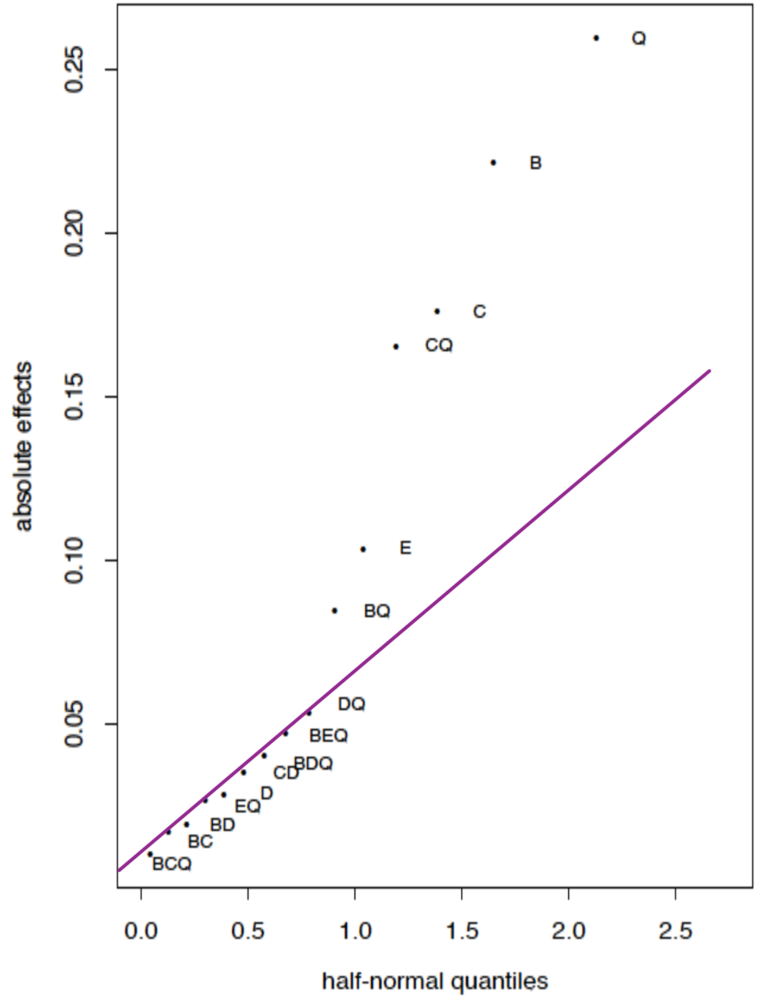
Suggest Significant Effects by Normal Probability Plot

- $\hat{y} = 7.6360 + 0.1106x_B + 0.0519x_E + 0.0881x_C - 0.1298x_Q + \cancel{0.0423x_Bx_Q} - 0.0827x_Cx_Q$





Effect	$\bar{y}$	$\ln s^2$
B	0.221	1.891
C	0.176	0.569
D	0.029	-0.247
E	0.104	0.216
Q	-0.260	0.280
BQ	0.085	-0.589
CQ	-0.165	0.598
DQ	0.054	1.111
EQ	0.027	0.129
BC	0.017	-0.002
BD	0.020	0.425
CD	-0.035	0.670
BCQ	0.010	-1.089
BDQ	-0.040	-0.432
BEQ	-0.047	0.854



Analysis for dispersion effects

Confirm Significant Effects by a Formal Test

- Recall that for the full factorial design $N = 2^k n$

$$\mathrm{Var(Effect)} = Var(2\hat{\beta}) = \frac{\sigma^2}{2^{k-2}n} = \frac{\sigma^2}{N2^{-2}}$$

$$N2^{-2} = 2^{k-p}2^{-2} = 2^{5-1-2}$$

  - for each effect, the estimate follows $N(0,\sigma^2/4)$
  - σ^2 can be estimated by s^2=0.017
- Thus, the significance of each effect can be confirmed by a t-test (with adjustment for multiple tests)
  - m = 15 if all effects are tested
  - m = 12 if 3-order interaction effects are ignored
- Advantage of the Normal-Probability-Plot method:
  - Can still work for experiments without replicates
- $\hat{y} = 7.6360 + 0.1106x_B + 0.0881x_C - 0.1298x_Q - 0.0827x_C x_Q$

Effect	$\bar{y}$	$\ln s^2$
B	0.221	1.891
C	0.176	0.569
D	0.029	-0.247
~~E~~	~~0.104~~	0.216
Q	-0.260	0.280
~~BQ~~	~~0.085~~	-0.589
CQ	-0.165	0.598
DQ	0.054	1.111
EQ	0.027	0.129
BC	0.017	-0.002
BD	0.020	0.425
CD	-0.035	0.670
BCQ	0.010	-1.089
BDQ	-0.040	-0.432
BEQ	-0.047	0.854

Fold-over Technique

- Suppose the original experiment is based on a 2_{III}^{7-4} design with generators

$$d_1: 4 = 12,\ 5 = 13,\ 6 = 23,\ 7 = 123$$

None of its main effects are clear

- To de-alias them, we can choose another 8 runs (no. 9-16->) with reversed signs for each of the 7 factors. This follow-up design d_2 has the generators

$$d_2: 4 = -12,\ 5 = -13, 6 = -23,\ 7 = 123$$

- With the extra degrees of freedom, we can introduce a new factor 8 (or a blocking variable) for run number 1-8, and −8 for run number 9-16
- The combined design $d_1 + d_2$ is a 2_{IV}^{8-4} design and thus all main effects are clear
- Its defining contrast subgroup is

$$\boxed{I = 1237 = 1256 = 1346 = 1457 = 2345 = 2467 = 3567}$$

Augmented Design Matrix Using Fold-Over Technique

	d_1							
Run	**1**	**2**	**3**	**4=12**	**5=13**	**6=23**	**7=123**	**8**
1	−	−	−	+	+	+	−	+
2	−	−	+	+	−	−	+	+
3	−	+	−	−	+	−	+	+
4	−	+	+	−	−	+	−	+
5	+	−	−	−	−	+	+	+
6	+	−	+	−	+	−	−	+
7	+	+	−	+	−	−	−	+
8	+	+	+	+	+	+	+	+
	d_2							
Run	**-1**	**-2**	**-3**	**-4**	**-5**	**-6**	**-7**	**-8**
9	+	+	+	−	−	−	+	−
10	+	+	−	−	+	+	−	−
11	+	−	+	+	−	+	−	−
12	+	−	−	+	+	−	+	−
13	−	+	+	+	+	−	−	−
14	−	+	−	+	−	+	+	−
15	−	−	+	−	+	+	+	−
16	−	−	−	−	−	−	−	−

Fold-over Technique: Version Two

- Suppose one factor, say 5, is very important. We want to de-alias 5 and all two-factor interactions involving 5
- Choose, instead, the following 2_{III}^{7-4} design

$$d_3: 4 = 12,\ 5 = -13,\ 6 = 23,\ 7 = 123$$

$d_1: 4 = 12,\ 5 = 13,\ 6 = 23,\ 7 = 123$

- Then the combined design $d_1 + d_3$ is a 2_{III}^{7-3} design with the generators

$$d': 4 = 12,\ 6 = 23,\ 7 = 123$$

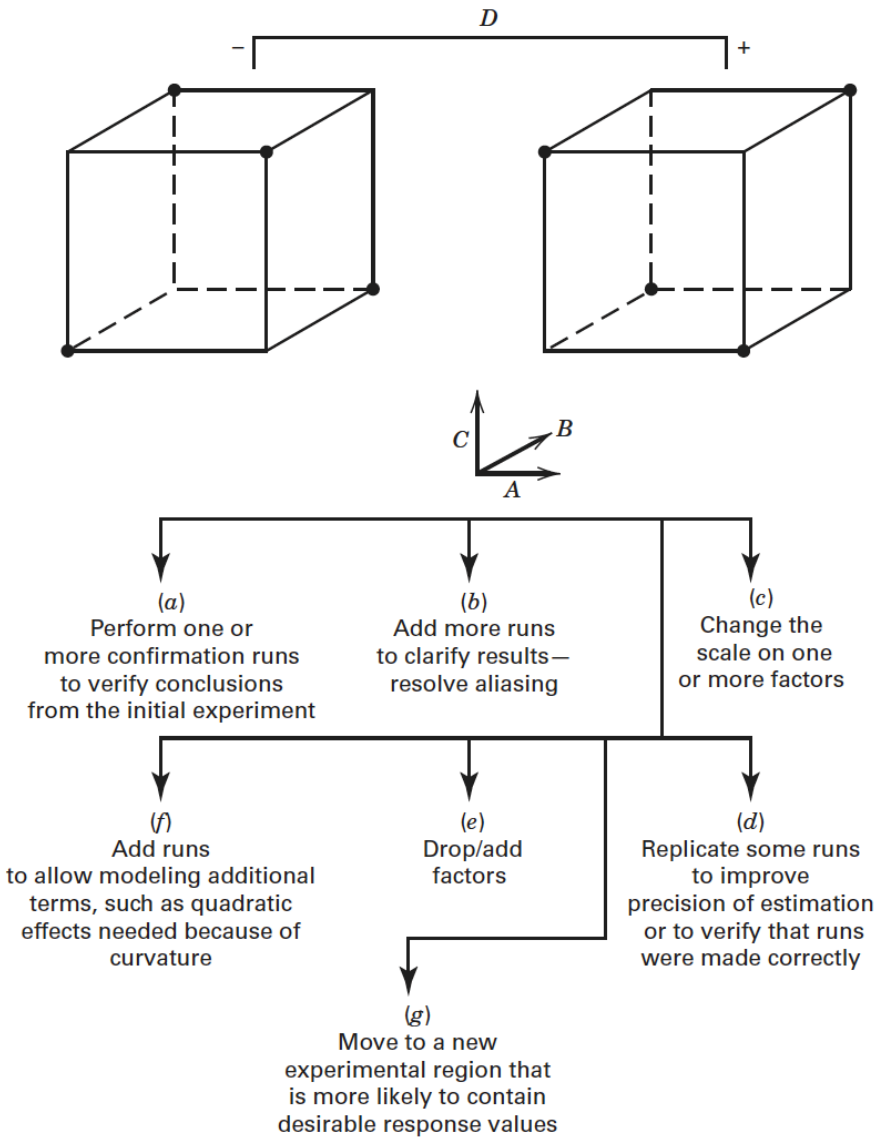
- Since 5 does not appear here, 5 is strongly clear and all two-factor interactions involving 5 are clear
- Choice between d_2 and d_3 depends on the priority given to the effects

Cons:
It requires doubling of the run size and can only de-alias a specific set of effects

Why do Fractional Factorial Designs Work?

- The **sparsity of effects** principle
  - There may be lots of factors, but few are important
  - System is dominated by main effects, low-order interactions
- The **projection** property
  - Every fractional factorial contains full factorials in fewer factors
- **Sequential** experimentation
  - Can add runs to a fractional factorial to resolve difficulties (or ambiguities) in interpretation
- Note: Ockham's razor in interpretation
  - A scientific principle that when one is confronted with several different possible interpretations of a phenomena, the simplest interpretation is usually the correct one

Possibilities for Follow-up Experimentation after an Initial Fractional Factorial Experiment



Plackett-Burman (PB) and Model Robust Screening Designs

- We looked at 2^{k-p} designs, which give us designs that have 8, 16, 32, 64, 128, etc. number of runs
- However, there is a pretty big gap between 16 and 32, 32 to 64, etc. We sometimes need other alternative designs besides these with a different number of observations
- A class of designs that allows us to create experiments with some number between these fractional factorial designs are the Plackett-Burman designs, for

  $$N = 12, [16], 20, 24, 28, [32], 36, 40, 44, 48, \ldots\ldots$$

  any number which is divisible by four
- These designs are similar to Resolution III designs, meaning you can estimate main effects clear of other main effects-often used for screening experiments (where the objective is to determine which factors from a list assembled by brainstorming are important enough to be studied in more detail in follow-up experiments)

- For run sizes that are powers of 2, they are the same as a 2^{k-p} fractional factorial design. For other run sizes, they retain the desirable orthogonality property of 2^{k-p} designs, but they do not have generators or a defining relation
- The designs for run sizes of 12, 20, and 24 can be created by cyclically rotating the factor levels for the first run

Run	A	B	C	D	E	F	G	H	J	K	L
1	+	+	−	+	+	+	−	−	−	+	−
2	−	+	+	−	+	+	+	−	−	−	+
3	+	−	+	+	−	+	+	+	−	−	−
4	−	+	−	+	+	−	+	+	+	−	−
5	−	−	+	+	+	+	−	+	+	+	−
6	−	−	−	+	−	+	+	−	+	+	+
7	+	−	−	−	+	−	+	+	−	+	+
8	+	+	−	−	−	+	−	+	+	−	+
9	+	+	+	−	−	−	+	−	+	+	−
10	−	+	+	+	−	−	−	+	−	+	+
11	+	−	+	+	+	−	−	−	+	−	+
12	−	−	−	−	−	−	−	−	−	−	−

Run Size	Factor Levels
12	+ + − + + + − − − + −
20	+ + − − + + + + − + − + − − − − + + −
24	+ + + + + − + − + + − − + + − − + − + − − − −

12-Run Plackett-Burman Design

Each combination of levels for any **pair** of factors appears the *same number of times*, throughout all the experimental runs

```
> library(FrF2)
> pb( nruns = 12, randomize=FALSE)#nfactors no more than nruns-1
```

```
> .libPaths()
```

Partial Confounding

- The cyclical pattern is a result of number theory properties that generate these orthogonal arrays. There is a lot of mathematical research behind these designs to achieve a matrix with orthogonal columns which is what we need
- Nongeometric designs: these designs cannot be represented as cubes
- Although some effects are orthogonal they do not have the same structure allowing complete or orthogonal correlation with the other two way and higher order interactions
- For 11 factors and 12 runs
  - The correlation matrix for main effects is identity
  - Every main effect is partially aliased with every two-factor interaction not involving itself

```
> design <- pb( nruns = 12)
> nmain <- apply(as.matrix(design), 2, as.numeric)
> cor(nmain)
> AB <- nmain[,1]*nmain[,2]
> BC <- nmain[,2]*nmain[,3]
> cor(AB, nmain[,1])  #0
> cor(AB, nmain[,3]) #-0.3333333
> cor(BC, nmain[,1])
```

- If you assume that interactions are not important, these are great designs, very efficient with small numbers of observations and useful
- If your assumption is wrong and there are interactions, it could show up as influencing one or the other main effects