

# 2<sup>k-p</sup> Fractional Factorial Designs

Raj Jain

Washington University in Saint Louis  
Saint Louis, MO 63130  
Jain@cse.wustl.edu

These slides are available on-line at:

<http://www.cse.wustl.edu/~jain/cse567-06/>



- 2<sup>k-p</sup> Fractional Factorial Designs
- Sign Table for a 2<sup>k-p</sup> Design
- Confounding
- Other Fractional Factorial Designs
- Algebra of Confounding
- Design Resolution

## 2<sup>k-p</sup> Fractional Factorial Designs

- Large number of factors
  - ⇒ large number of experiments
  - ⇒ full factorial design too expensive
  - ⇒ Use a fractional factorial design
- 2<sup>k-p</sup> design allows analyzing k factors with only 2<sup>k-p</sup> experiments.
  - 2<sup>k-1</sup> design requires only half as many experiments
  - 2<sup>k-2</sup> design requires only one quarter of the experiments

### Example: 2<sup>7-4</sup> Design

Expt No.	A	B	C	D	E	F	G
1	-1	-1	-1	1	1	1	-1
2	1	-1	-1	-1	-1	1	1
3	-1	1	-1	-1	1	-1	1
4	1	1	-1	1	-1	-1	-1
5	-1	-1	1	1	-1	-1	1
6	1	-1	1	-1	1	-1	-1
7	-1	1	1	-1	-1	1	-1
8	1	1	1	1	1	1	1

- Study 7 factors with only 8 experiments!

## Fractional Design Features

- Full factorial design is easy to analyze due to orthogonality of sign vectors.

Fractional factorial designs also use orthogonal vectors.

That is:

- The sum of each column is zero.

$$\sum_i x_{ij} = 0 \quad \forall j$$

$j$ th variable,  $i$ th experiment.

- The sum of the products of any two columns is zero.

$$\sum_i x_{ij} x_{il} = 0 \quad \forall j \neq l$$

- The sum of the squares of each column is  $2^{7-4}$ , that is, 8.

$$\sum_i x_{ij}^2 = 8 \quad \forall j$$

## Analysis of Fractional Factorial Designs

- **Model:**

$$y = q_0 + q_A x_A + q_B x_B + q_C x_C + q_D x_D + q_E x_E + q_F x_F + q_G x_G$$

- Effects can be computed using inner products.

$$\begin{aligned} q_A &= \sum_i y_i x_{Ai} \\ &= \frac{-y_1 + y_2 - y_3 + y_4 - y_5 + y_6 - y_7 + y_8}{8} \end{aligned}$$

$$\begin{aligned} q_B &= \sum_i y_i x_{Bi} \\ &= \frac{-y_1 - y_2 + y_3 + y_4 - y_5 - y_6 + y_7 + y_8}{8} \end{aligned}$$

## Example 19.1

I	A	B	C	D	E	F	G	y
1	-1	-1	-1	1	1	1	-1	20
1	1	-1	-1	-1	-1	1	1	35
1	-1	1	-1	-1	1	-1	1	7
1	1	1	-1	1	-1	-1	-1	42
1	-1	-1	1	1	-1	-1	1	36
1	1	-1	1	-1	1	-1	-1	50
1	-1	1	1	-1	-1	1	-1	45
1	1	1	1	1	1	1	1	82
317	101	35	109	43	1	47	3	Total
39.62	12.62	4.37	13.62	5.37	0.125	5.87	0.37	Total/8

- Factors A through G explain 37.26%, 4.74%, 43.40%, 6.75%, 0%, 8.06%, and 0.03% of variation, respectively.  
 ⇒ Use only factors C and A for further experimentation.

## Sign Table for a $2^{k-p}$ Design

Steps:

1. Prepare a sign table for a full factorial design with  $k-p$  factors.
2. Mark the first column I.
3. Mark the next  $k-p$  columns with the  $k-p$  factors.
4. Of the  $(2^{k-p}-k-p-1)$  columns on the right, choose  $p$  columns and mark them with the  $p$  factors which were not chosen in step 1.

## Example: $2^{7-4}$ Design

Expt No.	A	B	C	AB	AC	BC	ABC
1	-1	-1	-1	1	1	1	-1
2	1	-1	-1	-1	-1	1	1
3	-1	1	-1	-1	1	-1	1
4	1	1	-1	1	-1	-1	-1
5	-1	-1	1	1	-1	-1	1
6	1	-1	1	-1	1	-1	-1
7	-1	1	1	-1	-1	1	-1
8	1	1	1	1	1	1	1

## Example: $2^{4-1}$ Design

Expt No.	A	B	C	AB	AC	BC	D
1	-1	-1	-1	1	1	1	-1
2	1	-1	-1	-1	-1	1	1
3	-1	1	-1	-1	1	-1	1
4	1	1	-1	1	-1	-1	-1
5	-1	-1	1	1	-1	-1	1
6	1	-1	1	-1	1	-1	-1
7	-1	1	1	-1	-1	1	-1
8	1	1	1	1	1	1	1

## Confounding

- **Confounding:** Only the combined influence of two or more effects can be computed.

$$\begin{aligned}q_A &= \sum_i y_i x_{Ai} \\ &= \frac{-y_1 + y_2 - y_3 + y_4 - y_5 + y_6 - y_7 + y_8}{8}\end{aligned}$$

$$\begin{aligned}q_D &= \sum_i y_i x_{Di} \\ &= \frac{-y_1 + y_2 + y_3 - y_4 + y_5 - y_6 - y_7 + y_8}{8}\end{aligned}$$

## Confounding (Cont)

$$\begin{aligned}q_{ABC} &= \sum_i y_i x_{Ai} x_{Bi} x_{Ci} \\ &= \frac{-y_1 + y_2 + y_3 - y_4 + y_5 - y_6 - y_7 + y_8}{8}\end{aligned}$$

$$q_D = q_{ABC}$$

$$\begin{aligned}q_D + q_{ABC} &= \sum_i y_i x_{Ai} x_{Bi} x_{Ci} \\ &= \frac{-y_1 + y_2 + y_3 - y_4 + y_5 - y_6 - y_7 + y_8}{8}\end{aligned}$$

- $\Rightarrow$  Effects of D and ABC are confounded. Not a problem if  $q_{ABC}$  is negligible.

## Confounding (Cont)

- Confounding representation:  $D=ABC$

Other Confoundings:

$$\begin{aligned}
 q_A &= q_{BCD} = \sum_i y_i x_{Ai} \\
 &= \frac{-y_1 + y_2 - y_3 + y_4 - y_5 + y_6 - y_7 + y_8}{8}
 \end{aligned}$$

$$\Rightarrow A = BCD$$

$A=BCD$ ,  $B=ACD$ ,  $C=ABD$ ,  $AB=CD$ ,  $AC=BD$ ,  
 $BC=AD$ ,  $ABC=D$ , and  $I=ABCD$

- $I=ABCD \Rightarrow$  confounding of ABCD with the mean.

## Other Fractional Factorial Designs

- A fractional factorial design is not unique.  $2^p$  different designs.

Another  $2^{4-1}$  Experimental Design

Expt No.	A	B	C	D	AC	BC	ABC
1	-1	-1	-1	1	1	1	-1
2	1	-1	-1	-1	-1	1	1
3	-1	1	-1	-1	1	-1	1
4	1	1	-1	1	-1	-1	-1
5	-1	-1	1	1	-1	-1	1
6	1	-1	1	-1	1	-1	-1
7	-1	1	1	-1	-1	1	-1
8	1	1	1	1	1	1	1

- Confoundings:  $I=ABD$ ,  $A=BD$ ,  $B=AD$ ,  $C=ABCD$ ,  
 $D=AB$ ,  $AC=BCD$ ,  $BC=ACD$ ,  $ABC=CD$

Not as good as the previous design.

## Algebra of Confounding

- Given just one confounding, it is possible to list all other confoundings.
- Rules:
  - $I$  is treated as unity.
  - Any term with a power of 2 is erased.

$$I = ABCD$$

Multiplying both sides by A:

$$A = A^2BCD = BCD$$

Multiplying both sides by B, C, D, and AB:

## Algebra of Confounding (Cont)

$$B = AB^2CD = ACD$$

$$C = ABC^2D = ABD$$

$$D = ABCD^2 = ABC$$

$$AB = A^2B^2CD = CD$$

and so on.

- Generator polynomial:  $I=ABCD$

For the second design:  $I=ABC$ .

- In a  $2^{k-p}$  design,  $2^p$  effects are confounded together.

## Example 19.7

- In the  $2^{7-4}$  design:

$$D = AB, E = AC, F = BC, G = ABC$$

$$\Rightarrow I = ABD, I = ACE, I = BCF, I = ABCG$$

$$\Rightarrow I = ABD = ACE = BCF = ABCG$$

- Using products of all subsets:

$$I = ABD = ACE = BCF = ABCG = BCDE$$

$$= ACDF = CDG = ABEF = BEG$$

$$= AFG = DEF = ADEG = BDFG$$

$$= CEFG = ABCDEFG$$

## Example 19.7 (Cont)

- Other confoundings:

$$A = BD = CE = ABCF = BCG = ABCDE$$

$$= CDF = ACDG = BEF = ABEG$$

$$= FG = ADEF = DEG = ABDFG$$

$$= ACEFG = BCDEFG$$

## Design Resolution

- Order of an effect = Number of terms  
Order of  $ABCD = 4$ , order of  $I = 0$ .
- Order of a confounding = Sum of order of two terms  
E.g.,  $AB=CDE$  is of order 5.
- Resolution of a Design  
= Minimum of orders of confoundings
- Notation:  $R_{III} = \text{Resolution-III} = 2^{k-p}_{III}$
- Example 1:  $I=ABCD \Rightarrow R_{IV} = \text{Resolution-IV} = 2^{4-1}_{IV}$   
 $A=BCD, B=ACD, C=ABD, AB=CD, AC=BD,$   
 $BC=AD, ABC=D, \text{ and } I=ABCD$

## Design Resolution (Cont)

- Example 2:  
 $I = ABD \Rightarrow R_{III}$  design.
- Example 3:  
 $I = ABD = ACE = BCF = ABCG = BCDE$   
 $= ACDF = CDG = ABEF = BEG$   
 $= AFG = DEF = ADEG = BDFG$   
 $= ABDG = CEF = ABCDEFG$
- This is a resolution-III design.
- A design of higher resolution is considered a better design.

## Case Study 19.1: Latex vs. troff

Factors and Levels

	Factor	-Level	+Level
A	Program	Latex	troff-me
B	Bytes	2100	25000
C	Equations	0	10
D	Floats	0	10
E	Tables	0	10
F	Footnotes	0	10

## Case Study 19.1 (Cont)

□ Design:  $2^{6-1}$  with I=BCDEF

	Factor	Effect	% Variation
B	Bytes	12.0	39.4%
A	Program	9.4	24.4%
C	Equations	7.5	15.6%
AC	Program × Equations	7.2	14.4%
E	Tables	3.5	3.4%
F	Footnotes	1.6	0.70%

## Case Study 19.1: Conclusions

- Over 90% of the variation is due to: Bytes, Program, and Equations and a second order interaction.
- Text file size were significantly different making it's effect more than that of the programs.
- High percentage of variation explained by the "program  $\times$  Equation" interaction  
 $\Rightarrow$  Choice of the text formatting program depends upon the number of equations in the text. troff not as good for equations.

Program	CPU Time	
	# of Equations	
	-1(0)	1(10)
-1(Latex)	-9.7	-9.1
1(Troff)	-5.3	24.1

## Case Study 19.1: Conclusions (Cont)

- Low "Program  $\times$  Bytes" interaction  $\Rightarrow$  Changing the file size affects both programs in a similar manner.
- In next phase, reduce range of file sizes. Alternately, increase the number of levels of file sizes.

## Case Study 19.2: Scheduler Design

- Three classes of jobs: word processing, data processing, and background data processing.

Factors and Levels in the Scheduler Design Study

Symbol	Factor	Level -1	Level 1
A	Preemption	No	Yes
B	Time Slice	Small	Large
C	Queue Assignment	One Queue	Two Queues
D	Requeueing	Two Queues	Five Queues
E	Fairness	Off	On

- Design:  $2^{5-1}$  with  $I=ABCDE$

## Measured Throughputs

No.	A	B	C	D	E	$T_W$	$T_I$	$T_B$
1	-1	-1	-1	-1	1	15.0	25.0	15.2
2	1	-1	-1	-1	-1	11.0	41.0	3.0
3	-1	1	-1	-1	-1	25.0	36.0	21.0
4	1	1	-1	-1	1	10.0	15.7	8.6
5	-1	-1	1	-1	-1	14.0	63.9	7.5
6	1	-1	1	-1	1	10.0	13.2	7.5
7	-1	1	1	-1	1	28.0	36.3	20.2
8	1	1	1	-1	-1	11.0	23.0	3.0
9	-1	-1	-1	1	-1	14.0	66.1	6.4
10	1	-1	-1	1	1	10.0	9.1	8.4
11	-1	1	-1	1	1	27.0	34.6	15.7
12	1	1	-1	1	-1	11.0	23.0	3.0
13	-1	-1	1	1	1	14.0	26.0	12.0
14	1	-1	1	1	-1	11.0	38.0	2.0
15	-1	1	1	1	-1	25.0	35.0	17.2
16	1	1	1	1	1	11.0	22.0	2.0

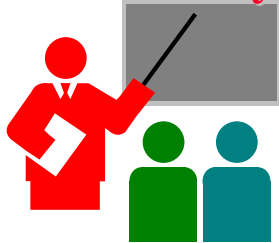
## Effects and Variation Explained

Confounded		$T_W$		$T_I$		$T_B$	
Effects		Esti- mate	Perc. Var.	Esti- mate	Perc. Var.	Esti- mate	Perc. Var.
1	2						
I	ABCDE	15.44		31.74		9.54	
A	BCDE	-4.81	55.5%	-8.62	31.0%	-4.86	58.8%
B	ACDE	3.06	22.5%	-3.54	5.2%	1.79	8.0%
C	ABDE	0.06	0.0%	0.43	0.1%	-0.62	1.0%
D	ABCE	-0.06	0.0%	-0.02	0.0%	-1.21	3.6%
AB	CDE	-2.94	20.7%	1.34	0.8%	-2.33	13.5%
AC	BDE	0.06	0.0%	0.49	0.1%	-0.44	0.5%
AD	BCE	0.19	0.1%	-0.08	0.0%	0.37	0.3%
BC	ADE	0.19	0.1%	0.44	0.1%	-0.12	0.0%
BD	ACE	0.06	0.0%	0.47	0.1%	-0.66	1.1%
CD	ABE	-0.19	0.1%	-1.91	1.5%	0.58	0.8%
DE	ABC	-0.06	0.0%	0.21	0.0%	-0.47	0.5%
CE	ABD	0.06	0.0%	1.21	0.6%	-0.16	0.1%
BE	ACD	0.31	0.2%	7.96	26.4%	-1.37	4.7%
AE	BCD	-0.56	0.8%	0.88	0.3%	0.28	0.2%
E	ABCD	0.19	0.1%	-9.01	33.8%	1.66	6.8%

## Case Study 19.2: Conclusions

- ❑ For word processing throughput ( $T_W$ ): A (Preemption), B (Time slice), and AB are important.
- ❑ For interactive jobs: E (Fairness), A (preemption), BE, and B (time slice).
- ❑ For background jobs: A (Preemption), AB, B (Time slice), E (Fairness).
- ❑ May use different policies for different classes of workloads.
- ❑ Factor C (queue assignment) or any of its interaction do not have any significant impact on the throughput.
- ❑ Factor D (Requiring) is not effective.
- ❑ Preemption (A) impacts all workloads significantly.
- ❑ Time slice (B) impacts less than preemption.
- ❑ Fairness (E) is important for interactive jobs and slightly important for background jobs.

## Summary



- ❑ Fractional factorial designs allow a large number of variables to be analyzed with a small number of experiments
- ❑ Many effects and interactions are confounded
- ❑ The resolution of a design is the sum of the order of confounded effects
- ❑ A design with higher resolution is considered better

## Exercise 19.1

Analyze the  $2^{4-1}$  design:

		$C_1$		$C_2$	
		$D_1$	$D_2$	$D_1$	$D_2$
$A_1$	$B_1$		40	15	
	$B_2$		20	10	
$A_2$	$B_1$	100			30
	$B_2$	120			50

- ❑ Quantify all main effects.
- ❑ Quantify percentages of variation explained.
- ❑ Sort the variables in the order of decreasing importance.
- ❑ List all confoundings.
- ❑ Can you propose a better design with the same number of experiments.
- ❑ What is the resolution of the design?

## Exercise 19.2

Is it possible to have a  $2^{4-1}_{III}$  design? a  $2^{4-1}_{II}$  design?  $2^{4-1}_{IV}$  design? If yes, give an example.

## Homework

- Updated Exercise 19.1  
Analyze the  $2^{4-1}$  design:

		$C_1$		$C_2$	
		$D_1$	$D_2$	$D_1$	$D_2$
$A_1$	$B_1$		30	15	
	$B_2$		20	10	
$A_2$	$B_1$	100			30
	$B_2$	110			50

- Quantify all main effects.
- Quantify percentages of variation explained.
- Sort the variables in the order of decreasing importance.
- List all confoundings.
- Can you propose a better design with the same number of experiments.
- What is the resolution of the design?