2^{k-p} Fractional Factorial Designs

Dr. John Mellor-Crummey

Department of Computer Science
Rice University

johnmc@cs.rice.edu

Goals for Today

Understand

- **2^{k-p} fractional factorial designs**
  - —preparing a sign table
  - —properties
  - —analysis
  - —confounding
  - —design resolution

2^{k-p} Fractional Factorial Designs

- **Motivation: full factorial design can be very expensive**
  - —large number of factors $\Rightarrow$ too many experiments
- **Pragmatic approach: 2^{k-p} fractional factorial designs**
  - —k factors
  - —2^{k-p} experiments
- **Fractional factorial design implications**
  - —2^{k-1} design $\Rightarrow$ half of the experiments of a full factorial design
  - —2^{k-2} design $\Rightarrow$ quarter of the experiments of a full factorial design

Example: Sign Table for a 2^{7-4} Design

Expt	I	A	B	C	D	E	F	G
1	1	-1	-1	-1	1	1	1	-1
2	1	1	-1	-1	-1	-1	1	1
3	1	-1	1	-1	-1	1	-1	1
4	1	1	1	-1	1	-1	-1	-1
5	1	-1	-1	1	1	-1	-1	1
6	1	1	-1	1	-1	1	-1	-1
7	1	-1	1	1	-1	-1	1	-1
8	1	1	1	1	1	1	1	1

- **Study the effects of 7 factors with 8 experiments**
- **Full factorial design would have required 2^7=128 experiments**

2^{k-p} Design Properties

Much like full factorial designs

Orthogonality of sign vectors ⇒ easy analysis

(Note: column I , the identity vector is not considered for 1, 2)

1. **Sum of each column in the sign table is 0**

   sum(S(:,j)) = 0, for all j > 1

2. **Sum of the product of any two columns is 0**

   sum(S(:,j) * S(:,k)) = 0, for all j, k > 1, j ≠ k

3. **Sum of the square of elements in any column is 2^{k-p}**

   sum(S(:,j) * S(:,j)) = 2^{k-p}, for all j

Model equation for 2^{7-4} design

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

Computing the Effects for a 2^{3-1} Design

Expt	I	A	B	C	y
1	1	-1	-1	1	y1
2	1	1	-1	-1	y2
3	1	-1	1	-1	y3
4	1	1	1	1	y4

$$y = q_0 + q_A x_A + q_B x_B + q_C x_C$$

Sign Table Equations

$$y1 = q_0 - q_A - q_B + q_C$$
$$y2 = q_0 + q_A - q_B - q_C$$
$$y3 = q_0 - q_A + q_B - q_C$$
$$y4 = q_0 + q_A + q_B + q_C$$

Solving for the Effects

$$q_0 = (y1 + y2 + y3 + y4)/4$$
$$q_1 = (-y1 + y2 - y3 + y4)/4$$
$$q_2 = (-y1 - y2 + y3 + y4)/4$$
$$q_3 = (y1 - y2 - y3 + y4)/4$$

In general: to solve for an effect, compute inner product of y column with factor sign table column and divide by 2^{k-p}

Example: a 2^{7-4} Design

- **Analyzing data with a 7-factor 2^{7-4} experimental design**

Expt	I	A	B	C	D	E	F	G	y
1	1	-1	-1	-1	1	1	1	-1	11
2	1	1	-1	-1	-1	-1	1	1	35
3	1	-1	1	-1	-1	1	-1	1	8
4	1	1	1	-1	1	-1	-1	-1	42
5	1	-1	-1	1	1	-1	-1	1	31
6	1	1	-1	1	-1	1	-1	-1	51
7	1	-1	1	1	-1	-1	1	-1	52
8	1	1	1	1	1	1	1	1	91
	321	117	65	129	29	1	57	9	Total
	40.1	14.63	8.125	16.13	3.625	0.125	7.125	1.125	Total/8

- **As before, effects of factors in last line of table**

- **Percent variation explained**

A	B	C	D	E	F	G
37.26	4.47	43.40	6.75	0.00	8.07	0.03

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

- **Prepare a sign table for a full factorial design with k-p factors**
  - —table of 2^{k-p} rows and columns
  - —first column with all 1's; mark it "`I`"
  - —next k-p columns: mark with chosen k-p factors
  - —of the 2^{k-p}-k+p-1 columns remaining, relabel p of them with remaining factors
- **Example: prepare a 2^{7-4} table**
  - —prepare a sign table for a 2^3 design for first 3 factors
  - —relabel rightmost 4 columns with additional factors D, E, F, G

I	A	B	C	D	E	F	G
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

Example: Prepare Sign Table for 2^{4-1} Design

- **Prepare a sign table for a 2^3 design for first 3 factors**
- **Arbitrarily pick one of the rightmost 4 columns: label it D**

I	A	B	C	AB	AC	BC	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

- **What effects can we compute with this design?**
  - **—main effects: q_A, q_B, q_C, q_D**
  - **—interactions: q_{AB}, q_{AC}, q_{BC}**

How Much Can We Cheat?

- Given k factors, what is the smallest possible sign table that can be used?
- Phrased another way, how large can p be?

Confounding

With 2^{k-p} fractional factorial designs, not all effects can be separated

- **Confounding: influence of 2 or more effects is inseparable**
- **Two or more effects are confounded if they use the same linear combination of responses**

I	A	B	C	AB	AC	BC	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

- **"Effect" D: $\sum_i y_i x_{Di}$ = $(-y_1+y_2+y_3-y_4+y_5-y_6-y_7+y_8)/8$**
- **"Effect" ABC: $\sum_i y_i x_{Ai} x_{Bi} x_{Ci}$ = $(-y_1+y_2+y_3-y_4+y_5-y_6-y_7+y_8)/8$**
- **In fact, $(-y_1+y_2+y_3-y_4+y_5-y_6-y_7+y_8)/8 = q_D + q_{ABC}$**
- **Impossible to separate D, ABC effects without 2^4 design**

Confounding Revealed

- **Is confounding a problem? Not necessarily!**
  - —interaction is small ⇒
    confounded effect is primarily the effect of factor
- **Consider preceding example**
  - —if q_{ABC} is small, then $q_{ABC} + q_D$ is approximately q_D
- **Other effects are confounded as well**
  - —every column in the design represents the sum of two effects
  - —four factors ⇒ 16 effects
  - —in a 2^{4-1} design, only 8 effects can be computed
  - —each quantity therefore represents the sum of two effects

A=BCD, B=ACD, C=ABD, AB=CD

AC=BD, BC=AD, ABC=D, I=ABCD

When does Confounding Occur

- 2^{k-p} fractional factorial design: 2^p effects confounded together
- Consider 2^{7-4} design
- From factor column re-labelings
  - —D=AB, E=AC, F=BC, G=ABC
- Implications
  - —I=ABD, I=ACE, I=BCF, I=ABCG

  and from those we derive …
  - —I=ABD=ACE=ACF=ABCG=BCDE=ACDF=CDG=ABEF=BEG=AFG= DEF=CEFG=ABCDEFG
- Some other confoundings
  - —A=BD=CE=CF=BCG=ABCDE=CDF=ACDG=BEF=ABEG=FG= ADEF=ACEFG=BCDEFG

Fractional Factorial Designs are Not Unique

- **One 2^{4-1} design**

I	A	B	C	AB	AC	BC	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

Which is better?

- **Another 2^{4-1} design**

I	A	B	C	D	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

Confoundings with higher-order interactions is better.

Algebra of Confounding

- Designs identified by confounding with `I`
  - e.g. two designs on previous slide: `I=ABCD`, `I=ABD`
  - confounding with `I` is called a *generator polynomial* for a design
    - `I=ABCD` represents confounding ABCD with mean (represented by `I`)
- Given one confounding, can generate all others using 2 rules
  1. Mean `I` is treated as unity, e.g. `IB = B`
  2. Any term with power 2 is erased, e.g. $A^2BC = BC$
- Example: consider confounding `I=ABCD`
- Multiplying both sides by A
  - $AI = A^2BCD \Rightarrow A = BCD$
- Similarly
  - $BI = AB^2CD \Rightarrow B = ACD$
  - $CI = ABC^2D \Rightarrow C = ABD$
  - $DI = ABCD^2 \Rightarrow D = ABC$
  - $ABI = A^2B^2CD \Rightarrow AB = CD$

Design Resolution

- **Measured by the order of effects that are confounded**
- **Order of an effect is the number of factors included**
  - **e.g. order(AB) = 2, order(I) = 0**
- **Definition: order of a confounding**
  - **i^{th} order effect confounded with j^{th} order effect ⇒ confounding is order i+j, e.g. order(AC=BCD) = 5**
- **Definition: design resolution = min. order of confounded effects**
- **Example**
  - **I=ABCD, A = BCD, B = ACD, C = ABD, D = ABC, AB = CD, AC=BD, BC=AD**
  - **design resolution = 4**
  - **design is said to be R_{IV}**
- **Notation: R_{IV} = Resolution-IV = 2_{IV}^{4-1}**

Design Resolution Examples

Design Resolution Examples for 2^{4-1} Designs

- **One 2^{4-1} design**

I	A	B	C	AB	AC	BC	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

I=ABCD

R_{IV} design

- **Another 2^{4-1} design**

I	A	B	C	D	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

I=ABD

R_{III} design

Case Study: Latex vs. Troff

	Factor	-1 Level	+1 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

- **Design: 2^{6-1} with I=BCDEF**

	Factor	Effect	% var
A	Program	9.4	24.4%
B	Bytes	12.0	39.4%
C	Equations	7.5	15.6%
AC	Program x equations	7.2	14.4%
E	Tables	03.5	3.4%
F	Footnotes	1.6	.70%

Latex vs. Troff Study Findings

- > 90% variation due to
  - —program, bytes, equations, and program x equations
- Effect of bytes of text was larger than effect of program
  - —file sizes were quite different
- High program x equations interaction
  - — troff is very slow for equations
- Low program x bytes interaction
  - —changing file size affects both similarly
- Directions for the next phase of study
  - —reduce range of file sizes or
  - —increase number of levels of file sizes