2^k Factorial Designs

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

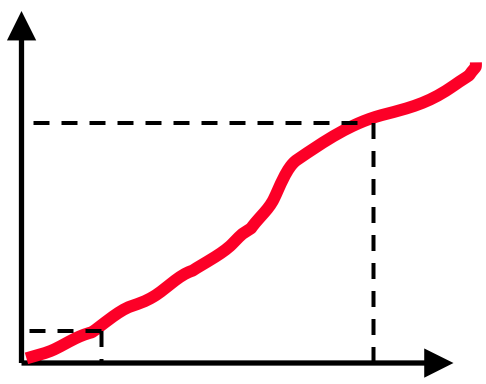
These slides are available on-line at:

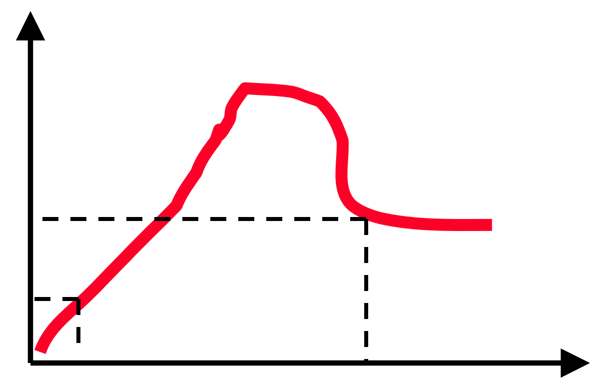
http://www.cse.wustl.edu/~jain/cse567-06/

- 2^2 Factorial Designs
- Model
- Computation of Effects
- Sign Table Method
- Allocation of Variation
- General 2^k Factorial Designs

2^k Factorial Designs

- k factors, each at two levels.
- Easy to analyze.
- Helps in sorting out impact of factors.
- Good at the beginning of a study.
- Valid only if the effect is unidirectional.
  E.g., memory size, the number of disk drives





2^2 Factorial Designs

- Two factors, each at two levels.

Performance in MIPS

Cache Size	Memory Size	
	4M Bytes	16M Bytes
1K	15	45
2K	25	75

$$x_A = \begin{vmatrix} -1 & \text{if 4M bytes memory} \\ 1 & \text{if 16M bytes memory} \end{vmatrix}$$

$$x_B = \begin{vmatrix} -1 & \text{if 1K bytes cache} \\ 1 & \text{if 2K bytes cache} \end{vmatrix}$$

Model

$$y = q_0 + q_A x_A + q_B x_B + q_{AB} x_A x_B$$

Observations:

$$15 = q_0 - q_A - q_B + q_{AB}$$

$$45 = q_0 + q_A - q_B - q_{AB}$$

$$25 = q_0 - q_A + q_B - q_{AB}$$

$$75 = q_0 + q_A + q_B + q_{AB}$$

Solution:

$$y = 40 + 20x_A + 10x_B + 5x_A x_B$$

Interpretation: Mean performance = 40 MIPS

Effect of memory = 20 MIPS; Effect of cache = 10 MIPS

Interaction between memory and cache = 5 MIPS.

Computation of Effects

Experiment	A	B	y
1	-1	-1	y_1
2	1	-1	y_2
3	-1	1	y_3
4	1	1	y_4

$$y = q_0 + q_A x_A + q_B x_B + q_{AB} x_A x_B$$

$$y_1 = q_0 - q_A - q_B + q_{AB}$$

$$y_2 = q_0 + q_A - q_B - q_{AB}$$

$$y_3 = q_0 - q_A + q_B - q_{AB}$$

$$y_4 = q_0 + q_A + q_B + q_{AB}$$

Computation of Effects (Cont)

Solution:

$$q_0 = \frac{1}{4}(y_1 + y_2 + y_3 + y_4)$$

$$q_A = \frac{1}{4}(-y_1 + y_2 - y_3 + y_4)$$

$$q_B = \frac{1}{4}(-y_1 - y_2 + y_3 + y_4)$$

$$q_{AB} = \frac{1}{4}(y_1 - y_2 - y_3 + y_4)$$

Notice that effects are linear combinations of responses.

Sum of the coefficients is zero $\Rightarrow$ **contrasts**.

Computation of Effects (Cont)

Experiment	A	B	y
1	-1	-1	y_1
2	1	-1	y_2
3	-1	1	y_3
4	1	1	y_4

$$q_A = \frac{1}{4}(-y_1 + y_2 - y_3 + y_4)$$

$$q_B = \frac{1}{4}(-y_1 - y_2 + y_3 + y_4)$$

Notice:

q_A = Column A × Column y

q_B = Column B × Column y

Sign Table Method

I	A	B	AB	y
1	-1	-1	1	15
1	1	-1	-1	45
1	-1	1	-1	25
1	1	1	1	75
160	80	40	20	Total
40	20	10	5	Total/4

Allocation of Variation

- Importance of a factor = proportion of the *variation* explained

$$\text{Sample Variance of } y = s_y^2 = \frac{\sum_{i=1}^{2^2}(y_i - \bar{y})^2}{2^2 - 1}$$

$$\text{Total Variation of } y = \text{SST} = \sum_{i=1}^{2^2}(y_i - \bar{y})^2$$

- For a 2^2 design:

$$\text{SST} = 2^2 q_A^2 + 2^2 q_B^2 + 2^2 q_{AB}^2 = \text{SSA} + \text{SSB} + \text{SSAB}$$

- Variation due to A = SSA = $2^2\, q_A{}^2$
- Variation due to B = SSB = $2^2\, q_B{}^2$
- Variation due to interaction = SSAB = $2^2\, q_{AB}{}^2$
- Fraction explained by A = $\frac{SSA}{SST}$

Variation ≠ Variance

Derivation

- Model:

$$y_i = q_0 + q_A x_{Ai} + q_B x_{Bi} + q_{AB} x_{Ai} x_{Bi}$$

Notice

1. The sum of entries in each column is zero:

$$\sum_{i=1}^4 x_{Ai} = 0; \sum_{i=1}^4 x_{Bi} = 0; \sum_{i=1}^4 x_{Ai} x_{Bi} = 0;$$

2. The sum of the squares of entries in each column is 4:

$$\begin{aligned} \sum_{i=1}^4 x_{Ai}^2 &= 4 \\ \sum_{i=1}^4 x_{Bi}^2 &= 4 \\ \sum_{i=1}^4 (x_{Ai} x_{Bi})^2 &= 4 \end{aligned}$$

Derivation (Cont)

3. The columns are orthogonal (inner product of any two columns is zero):

$$\sum_{i=1}^{4} x_{Ai} x_{Bi} = 0$$

$$\sum_{i=1}^{4} x_{Ai} \left(x_{Ai} x_{Bi} \right) = 0$$

$$\sum_{i=1}^{4} x_{Bi} \left(x_{Ai} x_{Bi} \right) = 0$$

Derivation (Cont)

- Sample mean $\bar{y}$

$$= \frac{1}{4}\sum_{i=1}^{4} y_i$$

$$= \frac{1}{4}\sum_{i=1}^{4} (q_0 + q_A x_{Ai} + q_B x_{Bi} + q_{AB} x_{Ai} x_{Bi})$$

$$= \frac{1}{4}\sum_{i=1}^{4} q_0 + \frac{1}{4} q_A \sum_{i=1}^{4} x_{Ai}$$

$$+ q_B \frac{1}{4}\sum_{i=1}^{4} x_{Bi} + q_{AB}\frac{1}{4}\sum_{i=1}^{4} x_{Ai} x_{Bi}$$

$$= q_0$$

Derivation (Cont)

- Variation of y

$$
\begin{aligned}
&= \sum_{i=1}^{4} (y_i - \bar{y})^2 \\
&= \sum_{i=1}^{4} (q_A x_{Ai} + q_B x_{Bi} + q_{AB} x_{Ai} x_{Bi})^2 \\
&= \sum_{i=1}^{4} (q_A x_{Ai})^2 + \sum_{i=1}^{4} (q_B x_{Bi})^2 \\
&+ \sum_{i=1}^{4} (q_{AB} x_{Ai} x_{Bi})^2 + \text{Product terms} \\
&= q_A^2 \sum_{i=1}^{4} (x_{Ai})^2 + q_B^2 \sum_{i=1}^{4} (x_{Bi})^2 \\
&+ q_{AB}^2 \sum_{i=1}^{4} (x_{Ai} x_{Bi})^2 + 0 \\
&= 4q_A^2 + 4q_B^2 + 4q_{AB}^2
\end{aligned}
$$

Example 17.2

- Memory-cache study:

$$\bar{y} = \frac{1}{4}(15 + 55 + 25 + 75) = 40$$

$$\begin{aligned} \text{Total Variation} &= \sum_{i=1}^{4} (y_i - \bar{y})^2 \\ &= (25^2 + 15^2 + 15^2 + 35^2) \\ &= 2100 \\ &= 4 \times 20^2 + 4 \times 10^2 + 4 \times 5^2 \end{aligned}$$

- Total variation= 2100

  Variation due to Memory = 1600 (76%)

  Variation due to cache = 400 (19%)

  Variation due to interaction = 100 (5%)

Case Study 17.1: Interconnection Nets

- Memory interconnection networks: Omega and Crossbar.
- Memory reference patterns: *Random* and *Matrix*
- Fixed factors:
  - Number of processors was fixed at 16.
  - Queued requests were not buffered but blocked.
  - Circuit switching instead of packet switching.
  - Random arbitration instead of round robin.
  - Infinite interleaving of memory $\Rightarrow$ no memory bank contention.

2^2 Design for Interconnection Networks

Factors Used in the Interconnection Network Study

Symbol	Factor	Level -1	Level 1
A	Type of the network	Crossbar	Omega
B	Address Pattern Used	Random	Matrix

A	B	Response: Throughput T	Response: 90% Transit N	Response: Response R
-1	-1	0.0641	3	1.655
1	-1	0.4220	5	2.378
-1	1	0.7922	2	1.262
1	1	0.4717	4	2.190

Interconnection Networks Results

Parameter	Mean Estimate			Variation Explained		
	T	N	R	T	N	R
q_0	0.5725	3.5	1.871			
q_A	0.0595	-0.5	-0.145	17.2%	20%	10.9%
q_B	-0.1257	1.0	0.413	77.0%	80%	87.8%
q_{AB}	-0.0346	0.0	0.051	5.8%	0%	1.3%

- Average throughput = 0.5725
- Most effective factor = B = Reference pattern
  $\Rightarrow$ The address patterns chosen are very different.
- Reference pattern explains $\mp$ 0.1257 (77%) of variation.
- Effect of network type = 0.0595
  Omega networks = Average + 0.0595
  Crossbar networks = Average - 0.0595
- Slight interaction (0.0346) between reference pattern and network type.

General 2^k Factorial Designs

- k factors at two levels each.

  2^k experiments.

  2^k effects:

  k main effects

  $\binom{k}{2}$ two factor interactions

  $\binom{k}{3}$ three factor interactions...

2^k Design Example

- Three factors in designing a machine:
  - Cache size
  - Memory size
  - Number of processors

	Factor	Level -1	Level 1
A	Memory Size	4MB	16MB
B	Cache Size	1kB	2kB
C	Number of Processors	1	2

2^k Design Example (cont)

Cache Size	4M Bytes		16M Bytes	
	1 Proc	2 Proc	1 Proc	2 Proc
1K Byte	14	46	22	58
2K Byte	10	50	34	86

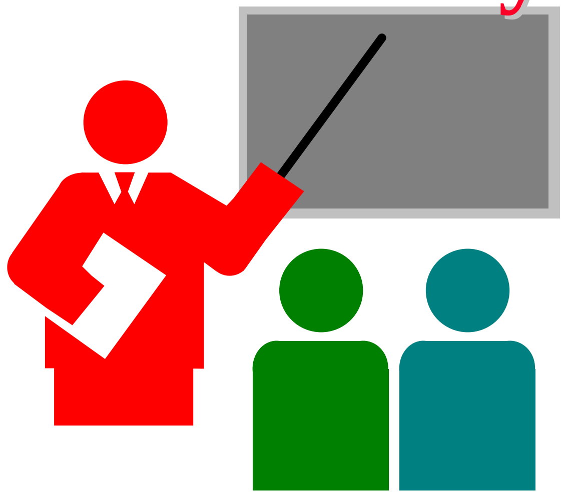
I	A	B	C	AB	AC	BC	ABC	y
1	-1	-1	-1	1	1	1	-1	14
1	1	-1	-1	-1	-1	1	1	22
1	-1	1	-1	-1	1	-1	1	10
1	1	1	-1	1	-1	-1	-1	34
1	-1	-1	1	1	-1	-1	1	46
1	1	-1	1	-1	1	-1	-1	58
1	-1	1	1	-1	-1	1	-1	50
1	1	1	1	1	1	1	1	86
320	80	40	160	40	16	24	9	Total
40	10	5	20	5	2	3	1	Total/8

Analysis of 2^k Design

$$
\begin{aligned}
\text{SST} &= 2^3(q_A^2 + q_B^2 + q_C^2 + q_{AB}^2 + q_{AC}^2 + q_{BC}^2 + q_{ABC}^2) \\
&= 8(10^2 + 5^2 + 20^2 + 5^2 + 2^2 + 3^2 + 1^2) \\
&= 800 + 200 + 3200 + 200 + 32 + 72 + 8 = 4512 \\
&= 18\% + 4\% + 71\% + 4\% + 1\% + 2\% + 0\% \\
&= 100\%
\end{aligned}
$$

- Number of Processors (C) is the most important factor.

Summary



- 2^k design allows k factors to be studied at two levels each
- Can compute main effects and all multi-factors interactions
- Easy computation using sign table method
- Easy allocation of variation using squares of effects

Exercise 17.1

Analyze the 2^3 design:

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

- Quantify main effects and all interactions.
- Quantify percentages of variation explained.
- Sort the variables in the order of decreasing importance.

Homework

Modified Exercise 17.1 Analyze the 2^3 design:

	A_1		A_2	
	C_1	C_2	C_1	C_2
B_1	110	15	120	10
B_2	60	30	40	50

- Quantify main effects and all interactions.
- Quantify percentages of variation explained.
- Sort the variables in the order of decreasing importance.