$2^k$ 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/cse567-17/
Overview

- $2^2$ 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
2² Factorial Designs

- Two factors, each at two levels.

<table>
<thead>
<tr>
<th>Cache Size</th>
<th>Performance in MIPS</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Memory Size</td>
</tr>
<tr>
<td></td>
<td>4M Bytes</td>
</tr>
<tr>
<td>1K</td>
<td>15</td>
</tr>
<tr>
<td>2K</td>
<td>25</td>
</tr>
</tbody>
</table>

\[ x_A = \begin{cases} 
-1 & \text{if 4M bytes memory} \\
1 & \text{if 16M bytes memory} 
\end{cases} \]

\[ x_B = \begin{cases} 
-1 & \text{if 1K bytes cache} \\
1 & \text{if 2K bytes cache} 
\end{cases} \]
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:

\[ y = 40 + 20 x_A + 10 x_B + 5 x_A x_B \]

**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

<table>
<thead>
<tr>
<th>Experiment</th>
<th>A</th>
<th>B</th>
<th>y</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>-1</td>
<td>-1</td>
<td>$y_1$</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>-1</td>
<td>$y_2$</td>
</tr>
<tr>
<td>3</td>
<td>-1</td>
<td>1</td>
<td>$y_3$</td>
</tr>
<tr>
<td>4</td>
<td>1</td>
<td>1</td>
<td>$y_4$</td>
</tr>
</tbody>
</table>

\[
y = q_0 + q_A x_A + q_B x_B + q_{AB} x_A x_B
\]

\[
y_1 = q_0 - q_A - q_B + q_{AB}
\]

\[
y_2 = q_0 + q_A - q_B - q_{AB}
\]

\[
y_3 = q_0 - q_A + q_B - q_{AB}
\]

\[
y_4 = q_0 + q_A + q_B + q_{AB}
\]
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)

<table>
<thead>
<tr>
<th>Experiment</th>
<th>A</th>
<th>B</th>
<th>y</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>-1</td>
<td>-1</td>
<td>$y_1$</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>-1</td>
<td>$y_2$</td>
</tr>
<tr>
<td>3</td>
<td>-1</td>
<td>1</td>
<td>$y_3$</td>
</tr>
<tr>
<td>4</td>
<td>1</td>
<td>1</td>
<td>$y_4$</td>
</tr>
</tbody>
</table>

\[
q_A = \frac{1}{4} \left(-y_1 + y_2 - y_3 + y_4\right)
\]

\[
q_B = \frac{1}{4} \left(-y_1 - y_2 + y_3 + y_4\right)
\]

**Notice:**

\[
q_A = \text{Column A} \not\equiv \text{Column } y
\]

\[
q_B = \text{Column B} \not\equiv \text{Column } y
\]
## Sign Table Method

<table>
<thead>
<tr>
<th>I</th>
<th>A</th>
<th>B</th>
<th>AB</th>
<th>y</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>-1</td>
<td>-1</td>
<td>1</td>
<td>15</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>-1</td>
<td>-1</td>
<td>45</td>
</tr>
<tr>
<td>1</td>
<td>-1</td>
<td>1</td>
<td>-1</td>
<td>25</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>75</td>
</tr>
<tr>
<td>160</td>
<td>80</td>
<td>40</td>
<td>20</td>
<td>Total</td>
</tr>
<tr>
<td>40</td>
<td>20</td>
<td>10</td>
<td>5</td>
<td>Total/4</td>
</tr>
</tbody>
</table>
Allocation of Variation

- Importance of a factor = proportion of the variation explained

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

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 = \( \frac{\text{SSA}}{\text{SST}} \) Variation \( \neq \) 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:
\[ \sum_{i=1}^{4} x_{Ai}^2 = 4 \]
\[ \sum_{i=1}^{4} x_{Bi}^2 = 4 \]
\[ \sum_{i=1}^{4} (x_{Ai} x_{Bi})^2 = 4 \]
Derivation (Cont)

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

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

\[
\sum_{i=1}^{4} x_{Ai} (x_{Ai} x_{Bi}) = 0
\]

\[
\sum_{i=1}^{4} x_{Bi} (x_{Ai} x_{Bi}) = 0
\]
Derivation (Cont)

- Sample mean $\bar{y}$

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

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

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

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

$$= q_0$$
Derivation (Cont)

- Variation of \( y \)

\[
= \sum_{i=1}^{4} (y_i - \bar{y})^2 \\
= \sum_{i=1}^{4} \left( q_A x_{A_i} + q_B x_{B_i} + q_{AB} x_{A_i} x_{B_i} \right)^2 \\
= \sum_{i=1}^{4} (q_A x_{A_i})^2 + \sum_{i=1}^{4} (q_B x_{B_i})^2 + \sum_{i=1}^{4} (q_{AB} x_{A_i} x_{B_i})^2 + \text{Product terms} \\
= q_A^2 \left( \sum_{i=1}^{4} (x_{A_i})^2 \right) + q_B^2 \left( \sum_{i=1}^{4} (x_{B_i})^2 \right) + q_{AB}^2 \left( \sum_{i=1}^{4} (x_{A_i} x_{B_i})^2 \right) + 0 \\
= 4q_A^2 + 4q_B^2 + 4q_{AB}^2
\]
Example 17.2

- Memory-cache study:

\[
\bar{y} = \frac{1}{4} (15 + 45 + 25 + 75) = 40
\]

\[
\text{Total Variation} = \sum_{i=1}^{4} (y_i - \bar{y})^2
\]

\[
= (25^2 + 5^2 + 15^2 + 35^2)
\]

\[
= 2100
\]

\[
= 4 \times 20^2 + 4 \times 10^2 + 4 \times 5^2
\]

- 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 ⇒ no memory bank contention.
2^2 Design for Interconnection Networks

Factors Used in the Interconnection Network Study

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Factor</th>
<th>Level</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>Type of the network</td>
<td>Crossbar Omega</td>
</tr>
<tr>
<td>B</td>
<td>Address Pattern Used</td>
<td>Random Matrix</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>A</th>
<th>B</th>
<th>Throughput T</th>
<th>90% Transit N</th>
<th>Response R</th>
</tr>
</thead>
<tbody>
<tr>
<td>-1</td>
<td>-1</td>
<td>0.6041</td>
<td>3</td>
<td>1.655</td>
</tr>
<tr>
<td>1</td>
<td>-1</td>
<td>0.7922</td>
<td>2</td>
<td>1.262</td>
</tr>
<tr>
<td>-1</td>
<td>1</td>
<td>0.4220</td>
<td>5</td>
<td>2.378</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>0.4717</td>
<td>4</td>
<td>2.190</td>
</tr>
</tbody>
</table>
Interconnection Networks Results

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Mean Estimate</th>
<th>Variation Explained</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>T</td>
<td>N</td>
</tr>
<tr>
<td>$q_0$</td>
<td>0.5725</td>
<td>3.5</td>
</tr>
<tr>
<td>$q_A$</td>
<td>0.0595</td>
<td>0.5</td>
</tr>
<tr>
<td>$q_B$</td>
<td>-0.1257</td>
<td>1.0</td>
</tr>
<tr>
<td>$q_{AB}$</td>
<td>-0.0346</td>
<td>0.0</td>
</tr>
</tbody>
</table>

- **Average throughput** = 0.5725
- **Most effective factor** = B = Reference pattern
  - 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

\[
y = q_0 + q_A x_A + q_B x_B + q_C x_C + q_{AB} x_A x_B + q_{AC} x_A x_C + q_{BC} x_B x_C + q_{ABC} x_A x_B x_C
\]
### 2^k Design Example (cont)

<table>
<thead>
<tr>
<th>Cache Size</th>
<th>4M Bytes</th>
<th></th>
<th></th>
<th>16M Bytes</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1 Proc</td>
<td>2 Proc</td>
<td></td>
<td>1 Proc</td>
<td>2 Proc</td>
<td></td>
</tr>
<tr>
<td>1K Byte</td>
<td>14</td>
<td>46</td>
<td></td>
<td>22</td>
<td>58</td>
<td></td>
</tr>
<tr>
<td>2K Byte</td>
<td>10</td>
<td>50</td>
<td></td>
<td>34</td>
<td>86</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>I</th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>AB</th>
<th>AC</th>
<th>BC</th>
<th>ABC</th>
<th>y</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>-1</td>
<td>-1</td>
<td>-1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>-1</td>
<td>14</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>-1</td>
<td>-1</td>
<td>-1</td>
<td>-1</td>
<td>1</td>
<td>1</td>
<td>22</td>
</tr>
<tr>
<td>1</td>
<td>-1</td>
<td>1</td>
<td>-1</td>
<td>-1</td>
<td>1</td>
<td>-1</td>
<td>1</td>
<td>10</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>-1</td>
<td>1</td>
<td>-1</td>
<td>-1</td>
<td>-1</td>
<td>34</td>
</tr>
<tr>
<td>1</td>
<td>-1</td>
<td>-1</td>
<td>1</td>
<td>1</td>
<td>-1</td>
<td>-1</td>
<td>1</td>
<td>46</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>-1</td>
<td>1</td>
<td>-1</td>
<td>1</td>
<td>-1</td>
<td>-1</td>
<td>58</td>
</tr>
<tr>
<td>1</td>
<td>-1</td>
<td>1</td>
<td>1</td>
<td>-1</td>
<td>-1</td>
<td>1</td>
<td>-1</td>
<td>50</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>86</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>320</th>
<th>80</th>
<th>40</th>
<th>160</th>
<th>40</th>
<th>16</th>
<th>24</th>
<th>8</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>40</td>
<td>10</td>
<td>5</td>
<td>20</td>
<td>5</td>
<td>2</td>
<td>3</td>
<td>1</td>
<td>Total/8</td>
</tr>
</tbody>
</table>

Washington University in St. Louis

http://www.cse.wustl.edu/~jain/cse567-17/

©2017 Raj Jain
Analysis of $2^k$ Design

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

- 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:

<table>
<thead>
<tr>
<th></th>
<th>$A_1$</th>
<th></th>
<th>$A_2$</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$C_1$</td>
<td>$C_2$</td>
<td>$C_1$</td>
<td>$C_2$</td>
</tr>
<tr>
<td>$B_1$</td>
<td>100</td>
<td>15</td>
<td>120</td>
<td>10</td>
</tr>
<tr>
<td>$B_2$</td>
<td>40</td>
<td>30</td>
<td>20</td>
<td>50</td>
</tr>
</tbody>
</table>

- Quantify main effects and all interactions.
- Quantify percentages of variation explained.
- Sort the variables in the order of decreasing importance.
Modified Exercise 17.1 Analyze the $2^3$ design:

<table>
<thead>
<tr>
<th></th>
<th>$A_1$</th>
<th></th>
<th>$A_2$</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$C_1$</td>
<td>$C_2$</td>
<td>$C_1$</td>
<td>$C_2$</td>
</tr>
<tr>
<td>$B_1$</td>
<td>110</td>
<td>15</td>
<td>120</td>
<td>10</td>
</tr>
<tr>
<td>$B_2$</td>
<td>60</td>
<td>30</td>
<td>40</td>
<td>50</td>
</tr>
</tbody>
</table>

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

CSE567M: Computer Systems Analysis (Spring 2013),
https://www.youtube.com/playlist?list=PLjGG94etKypJEKjNAa1n_1X0bWWNyZcof

CSE473S: Introduction to Computer Networks (Fall 2011),
https://www.youtube.com/playlist?list=PLjGG94etKypJWO5PMh8Azcg5e_10TiDw

Wireless and Mobile Networking (Spring 2016),
https://www.youtube.com/playlist?list=PLjGG94etKypKeb0nzyN9tSs_HCd5c4wXF

CSE571S: Network Security (Fall 2011),
https://www.youtube.com/playlist?list=PLjGG94etKypKvzfVtutHcPFJXumyy93u

Video Podcasts of Prof. Raj Jain's Lectures,
https://www.youtube.com/channel/UCN4-5wzNP9-ruOzQMs-8NUw