2^k Factorial Designs and 2^kr Designs with Replications

Dr. John Mellor-Crummey

Department of Computer Science
Rice University

johnmc@cs.rice.edu
Goals for Today

Understand

• $2^k$ factorial designs
  — sign table method
  — properties
  — analysis
  — allocating variation

• $2^k r$ designs with replications
  — error estimation
  — confidence intervals for effects

• Multiplicative models
2^k Factorial Designs

• Determine effects of k factors, each at two levels
• Total of 2^k experiments required
• Total of 2^k effects
  — k main effects
  — \binom{k}{2} two-factor interactions
  — \binom{k}{3} three-factor interactions
  — \binom{k}{j} j factor interactions, j \leq k
Example: a 2^3 Design

Evaluate impact of memory, cache size, # processors

<table>
<thead>
<tr>
<th>Factor</th>
<th>Level -1</th>
<th>Level 1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Memory Size, A</td>
<td>1GB</td>
<td>2GB</td>
</tr>
<tr>
<td>Cache Size, B</td>
<td>4MB</td>
<td>6MB</td>
</tr>
<tr>
<td># processors, C</td>
<td>1</td>
<td>2</td>
</tr>
</tbody>
</table>

Model performance using a non-linear equation in $x_A$ and $x_B$

$$y = q_0 + q_A x_A + q_B x_B + q_{AB} x_A x_B + q_{AC} x_A x_C +$$
$$q_{BC} x_B x_C + q_{ABC} x_A x_B x_C$$

Solve using sign table method
Calculating Effects with a Sign 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>Total</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>
</tr>
</thead>
<tbody>
<tr>
<td>Total/8</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>
</tr>
</tbody>
</table>

$q_0$ $q_A$ $q_B$ $q_C$ $q_{AB}$ $q_{AC}$ $q_{BC}$ $q_{ABC}$
Calculating Effects

\[
SST = 2^3 \left( q_A^2 + q_B^2 + q_C^2 + q_{AB}^2 + q_{AC}^2 + q_{BC}^2 + q_{ABC}^2 \right)
\]

\[
SST = 2^3 (10^2 + 5^2 + 20^2 + 5^2 + 2^2 + 3^2 + 1^2) = 4512
\]

<table>
<thead>
<tr>
<th>40</th>
<th>10</th>
<th>5</th>
<th>20</th>
<th>5</th>
<th>2</th>
<th>3</th>
<th>1</th>
</tr>
</thead>
<tbody>
<tr>
<td>q_0</td>
<td>q_A</td>
<td>q_B</td>
<td>q_C</td>
<td>q_{AB}</td>
<td>q_{AC}</td>
<td>q_{BC}</td>
<td>q_{ABC}</td>
</tr>
</tbody>
</table>

Effects

— SSA/SST \quad = \quad 2^3(10^2)/4512 \quad = \quad .18 = 18%
— SSB/SST \quad = \quad 2^3(5^2)/4512 \quad = \quad .04 = 4%
— SSC/SST \quad = \quad 2^3(20^2)/4512 \quad = \quad .71 = 71%
— SSAB/SST \quad = \quad 2^3(5^2)/4512 \quad = \quad .04 = 4%
— SSAC/SST \quad = \quad 2^3(2^2)/4512 \quad = \quad .01 = 1%
— SSBC/SST \quad = \quad 2^3(3^2)/4512 \quad = \quad .02 = 2%
— SSABC/SST \quad = \quad 2^3(1^2)/4512 \quad = \quad .00 = 0%
$2^k r$ Factorial Designs with Replications
2^{kr} Factorial Designs

- Cannot estimate errors with \(2^k\) factorial design
  — no experiment is repeated
- To quantify experimental errors
  — repeat measurements with same factor combinations
  — analyze using sign table
- \(2^{kr}\) design
  — \(r\) replications of \(2^k\) experiments (each of \(2^k\) factor combinations)
- Model

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

(e = experimental error)
Computing Effects for a $2^2r$ Design

<table>
<thead>
<tr>
<th>I</th>
<th>A</th>
<th>B</th>
<th>AB</th>
<th>$\bar{y}_i.$</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>48</td>
</tr>
<tr>
<td>1</td>
<td>-1</td>
<td>1</td>
<td>-1</td>
<td>24</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>77</td>
</tr>
<tr>
<td>164</td>
<td>86</td>
<td>38</td>
<td>20</td>
<td>total</td>
</tr>
<tr>
<td>41</td>
<td>21.5</td>
<td>9.5</td>
<td>5</td>
<td>total/4</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$y_{ij}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>15</td>
</tr>
<tr>
<td>45</td>
</tr>
<tr>
<td>25</td>
</tr>
<tr>
<td>75</td>
</tr>
</tbody>
</table>

- Factor combination = sign table row
- For each factor combination $i$, $r$ replicants (here, $r = 3$)
- $\bar{y}_i.$ = mean observation for $i^{th}$ factor combination
  $$\bar{y}_i. = \frac{1}{r} \sum y_{ij}$$
- Solve for $q_i$ coefficients using sign table and $\bar{y}_i.$ values
Estimating Experimental Errors for a $2^2r$ Design

• Model predicts following value for factor combination $i$

$$\hat{y}_i = q_0 + q_A x_{Ai} + q_B x_{Bi} + q_{AB} x_{Ai} x_{Bi}$$

• Errors

$$e_{ij} = y_{ij} - \hat{y}_i$$

$$= y_{ij} - q_0 + q_A x_{Ai} + q_B x_{Bi} + q_{AB} x_{Ai} x_{Bi}$$

$$= y_{ij} - \bar{y}_i.$$

• Total error across row is 0 by design
• Corollary: total error (across all rows) is also 0
### Sum of Squared Errors for a $2^23$ Design

$$SSE = \sum_{i=1}^{2^2} \sum_{j=1}^r e_{ij}$$

$$SSE = 0^2 + 3^2 + (-3)^2 + (-3)^2 + 0^2 + 3^2 + 1^2 + 4^2 + (-5)^2 + (-2)^2 + (-2)^2 + 4^2 = 102$$

<table>
<thead>
<tr>
<th>I</th>
<th>A</th>
<th>B</th>
<th>AB</th>
<th>$\bar{y}_i$</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>48</td>
</tr>
<tr>
<td>1</td>
<td>-1</td>
<td>1</td>
<td>-1</td>
<td>24</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>77</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$y_{ij}$</th>
<th>$e_{ij}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>15 18 12</td>
<td>0 3 -3</td>
</tr>
<tr>
<td>45 48 51</td>
<td>-3 0 3</td>
</tr>
<tr>
<td>25 28 19</td>
<td>1 4 -5</td>
</tr>
<tr>
<td>75 75 81</td>
<td>-2 -2 4</td>
</tr>
</tbody>
</table>
Mean Response for a $2^2r$ Design

- Model for $2^2r$ design

$$y_{ij} = q_0 + q_A x_{Ai} + q_B x_{Bi} + q_{AB} x_{Ai} x_{Bi} + e_{ij}$$

- Sum across all $2^2r$ observations =

$$\sum_{i,j} y_{ij} = \sum_{i,j} q_0 + \sum_{i,j} q_A x_{Ai} + \sum_{i,j} q_B x_{Bi} + \sum_{i,j} q_{AB} x_{Ai} x_{Bi} + \sum_{i,j} e_{ij}$$

- Since sum of x’s, their products and errors are all 0

$$\sum_{i,j} y_{ij} = \sum_{i,j} q_0 = 2^2 r q_0$$

mean response

$$\bar{y} = \frac{1}{2^2 r} \sum_{i,j} y_{ij} = \frac{1}{2^2 r} 2^2 r q_0 = q_0$$
SST and SSE for a $2^2 r$ Design

$$SST = \sum_{i,j} \left( y_{ij} - \bar{y}_{..} \right)^2 = \sum_{i,j} y_{ij}^2 - \sum_{i,j} \bar{y}_{..}^2$$

$$SSY = \sum_{i,j} y_{ij}^2 = \sum_{i,j} q_0^2 + \sum_{i,j} q_A^2 x_{Ai}^2 + \sum_{i,j} q_B^2 x_{Bi}^2 + \sum_{i,j} q_{AB}^2 x_{Ai}^2 x_{Bi}^2 + \sum_{i,j} e_{ij}^2 + \text{product terms (0 by design of sign table)}$$

$$SSY = SS0 + SSA + SSB + SSAB + SSE$$

$$SST = SSY - SS0 = SSA + SSB + SSAB + SSE$$

$$SSE = SSY - SS0 - SSA - SSB - SSAB$$

$$SSE = SSY - 2^2 r(q_0^2 + q_A^2 + q_B^2 + q_{AB}^2)$$

$$SS0 = \sum_{i,j} \bar{y}_{..}^2 = \sum_{i,j} q_0^2$$
Allocating Variation for a $2^2r$ Design

- Fraction of variation attributed to A = $\frac{SSA}{SST}$
- Fraction of variation attributed to B = $\frac{SSB}{SST}$
- Fraction of variation attributed to interaction of A & B = $\frac{SSAB}{SST}$
- Fraction of variation attributed to error = $\frac{SSE}{SST}$
Variance for $q_0$

- Effects computed from a sample are random variables
- Confidence intervals for effects
  — compute from variance of sample estimates
- If errors $\sim N(0, \sigma_e^2)$, model implies that $y \sim N(\bar{y}, \sigma_e^2)$
- Consider effect $q_0$
  \[ q_0 = 1 \frac{1}{2^2 r} \sum_{i,j} y_{ij} \]
  — $q_0$ = linear combination of normally distributed variables
  — therefore, $q_0$ is normally distributed as well
- Variance of $q_0$
  \[ \text{var}(q_0) = \frac{\sigma_e^2}{2^2 r} \]
Confidence Intervals for Effects in $2^2r$ Designs

• Estimate variance of errors from SSE

$$s_e^2 = \frac{SSE}{2^2(r-1)} = \text{MSE}$$

• Denominator = $2^2(r-1)$ = DOF = # independent terms in SSE
  —error terms for r replications all add to 0
  —only r-1 are independent

• Std dev of effects

$$S_{q_0} = S_{q_A} = S_{q_B} = S_{q_{AB}} = \frac{s_e}{\sqrt{2^2 r}}$$

• Confidence intervals for effects

$$q_i \pm t_{[1-\alpha/2;2^2(r-1)]} S_{q_i}$$

• Effect is significant if its confidence interval does not contain 0
Returning to our Example …

Memory/Cache Study

- Standard deviation of errors
  \[ s_e = \sqrt{\frac{SSE}{2^2(r-1)}} = \sqrt{\frac{102}{8}} = \sqrt{12.75} = 3.57 \]

- Standard deviation of effects
  \[ s_{q_i} = s_e / \sqrt{2^2 r} = 3.57 / \sqrt{12} = 1.03 \]

- Confidence intervals for effects
  — 90% confidence: \[ t_{[1-\alpha/2; 2^2(r-1)]} = t_{[.95; 8]} = 1.86 \]
  \[ q_i \pm t_{[1-\alpha/2; 2^2(r-1)]} s_{q_i} = q_i \pm (1.86)(1.03) = q_i \pm 1.92 \]
  \[ q_0 = (30.08, 42.91), \ q_A = (19.58, 23.41), \ q_B = (7.58, 11.41), \ q_{AB} = (3.08, 6.91) \]

- No confidence intervals include 0 \(\Rightarrow\) all significant
Assumptions

• Model errors are statistically independent
• Errors are additive
• Errors are normally distributed
• Errors have constant standard deviation $\sigma_e$
• Effects of factors are additive
• observations are independent and normally distributed with constant variance
Visual Tests for Verifying Assumptions

• Independent errors
  — scatter plot of residuals vs. predicted responses
    – should not have any trend
    – magnitude residuals < magnitude responses/10 ⇒ ignore trends
  — plot residuals as a function of experiment number
    – trend up or down implies other factors or side effects

• Normally distributed errors
  — normal quantile-quantile plot of errors should be linear

• Constant standard deviation of errors
  — scatter plot of $y$ for various levels of factors
  — if spread at one level significantly different than at another
    – need transformation

• What to do
  — if any test fails or $y_{max}/y_{min}$ is large, investigate multiplicative model
Multiplicative Models for $2^2r$ Experiments

- **Additive model**
  \[ y_{ij} = q_0 + q_A x_{Ai} + q_B x_{Bi} + q_{AB} x_{Ai} x_{Bi} + e_{ij} \]

- **Not applicable if effects don’t add**
  — example: execution time of workloads
    - factor A: $i^{th}$ processor speed = $v_i$ instructions/sec
    - factor B: $j^{th}$ workload size = $w_j$ instructions
  — the two effects multiply: execution time $y_{ij} = v_i \times w_j$

- **Use logarithm to produce additive model**
  — $\log(y_{ij}) = \log(v_i) + \log(w_j)$

- **Better model**
  \[ y'_{ij} = q_0 + q_A x_{Ai} + q_B x_{Bi} + q_{AB} x_{Ai} x_{Bi} + e_{ij} \]
  where \[ y'_{ij} = \log(y_{ij}) \]
Interpreting a Multiplicative Model

\[ \log(y) = q_0 + q_A x_A + q_B x_B + q_{AB} x_A x_B + e \]

\[ y = 10^{q_0} 10^{q_A x_A} 10^{q_B x_B} 10^{q_{AB} x_A x_B} 10^e \]

\[ u_A = 10^{q_A}, u_B = 10^{q_B}, u_{AB} = 10^{q_{AB}}, \]

- \( u_A \) is ratio of MIPS rating of two processors
- \( u_B \) ratio of size of the two workloads
- Antilog of additive mean \( q_0 \) yields geometric mean

\[ \hat{y} = 10^{q_0} = (y_1 y_2 \ldots y_n)^{1/n} \]

where \( n = 2^r \)
Example: Execution Times

- Analysis using an additive model
  
  factor A = proc speed; factor B = instructions

<table>
<thead>
<tr>
<th>I</th>
<th>A</th>
<th>B</th>
<th>AB</th>
<th>( \bar{y}_i )</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>-1</td>
<td>-1</td>
<td>1</td>
<td>104.170</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>-1</td>
<td>-1</td>
<td>1.003</td>
</tr>
<tr>
<td>1</td>
<td>-1</td>
<td>1</td>
<td>-1</td>
<td>1.003</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0.013</td>
</tr>
<tr>
<td></td>
<td>106.19</td>
<td>-104.15</td>
<td>-104.15</td>
<td>102.17</td>
</tr>
<tr>
<td></td>
<td>26.55</td>
<td>-26.04</td>
<td>-26.04</td>
<td>25.54</td>
</tr>
</tbody>
</table>

\( q_0 \) \quad \( q_A \) \quad \( q_B \) \quad \( q_{AB} \)

\( y_{ij} \)

\[
\begin{array}{ccc}
85.10 & 79.50 & 147.90 \\
.891  & 1.047 & 1.072 \\
.955  & .933  & 1.122 \\
.0148 & .0126 & .0118 \\
\end{array}
\]
Visual Tests for Additive Model
Why is This Model Invalid?

- Physical consideration
  - workload & processor effects don’t add

- Large range for $y$
  - $y_{\text{max}} / y_{\text{min}} = 147.90 / 0.0118 = 12,534 \Rightarrow \text{log transform}$
    - taking arithmetic mean of 104.17 and 0.013 is inappropriate

- Residuals are not small compared to response

- Spread of residuals is larger at larger value of response
  - suggests log transformation
Analysis using a Multiplicative Model

Data after log transformation
factor A = proc speed; factor B = instructions

<table>
<thead>
<tr>
<th>I</th>
<th>A</th>
<th>B</th>
<th>AB</th>
<th>$\bar{y}_i.$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>-1</td>
<td>-1</td>
<td>1</td>
<td>2.00</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>-1</td>
<td>-1</td>
<td>0.00</td>
</tr>
<tr>
<td>1</td>
<td>-1</td>
<td>1</td>
<td>-1</td>
<td>0.00</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>-1.89</td>
</tr>
<tr>
<td>.11</td>
<td>-3.89</td>
<td>-3.89</td>
<td>.11</td>
<td>total</td>
</tr>
<tr>
<td>.03</td>
<td>-0.97</td>
<td>-0.97</td>
<td>.03</td>
<td>total/4</td>
</tr>
</tbody>
</table>

$q_0$  $q_A$  $q_B$  $q_{AB}$

$y_{ij}$

<table>
<thead>
<tr>
<th></th>
<th>1.93</th>
<th>1.90</th>
<th>2.17</th>
</tr>
</thead>
<tbody>
<tr>
<td>-0.05</td>
<td>.02</td>
<td>.03</td>
<td></td>
</tr>
<tr>
<td>-0.02</td>
<td>-0.03</td>
<td>.05</td>
<td></td>
</tr>
<tr>
<td>-1.83</td>
<td>-1.90</td>
<td>-1.93</td>
<td></td>
</tr>
</tbody>
</table>
Visual Tests for Multiplicative Model

- Conclusion: Multiplicative model is better than additive model
## Variation Explained by the Models

<table>
<thead>
<tr>
<th>Factor</th>
<th>Effect</th>
<th>%var</th>
<th>Conf Interval</th>
<th>Effect</th>
<th>%var</th>
<th>Conf Interval</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>26.55</td>
<td></td>
<td>(16.35,36.74)</td>
<td>.03</td>
<td>0.0%</td>
<td>(16.35,36.74)</td>
</tr>
<tr>
<td>A</td>
<td>-26.04</td>
<td>30.1%</td>
<td>(-36.23,-15.84)</td>
<td>-.97</td>
<td>49.9%</td>
<td>(-1.02,-.93)</td>
</tr>
<tr>
<td>B</td>
<td>-26.04</td>
<td>30.1%</td>
<td>(-36.23,-15.84)</td>
<td>-.97</td>
<td>49.9%</td>
<td>(-1.02,-.93)</td>
</tr>
<tr>
<td>AB</td>
<td>25.54</td>
<td>29.0%</td>
<td>(15.35,35.74)</td>
<td>.03</td>
<td>0.0%</td>
<td>(-0.02,.07)*</td>
</tr>
<tr>
<td>e</td>
<td>10.8%</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

* = not significant

- With multiplicative model: interaction almost 0
- Little unexplained variation: .2%