Two Factors
Full Factorial Design
without Replications

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/cse567-11/
Overview

- Computation of Effects
- Estimating Experimental Errors
- Allocation of Variation
- ANOVA Table
- Visual Tests
- Confidence Intervals For Effects
- Multiplicative Models
- Missing Observations
Two Factors Full Factorial Design

- Used when there are two parameters that are carefully controlled
- Examples:
  - To compare several processors using several workloads.
  - To determining two configuration parameters, such as cache and memory sizes
- Assumes that the factors are categorical. For quantitative factors, use a regression model.
- A full factorial design with two factors $A$ and $B$ having $a$ and $b$ levels requires $ab$ experiments.
- First consider the case where each experiment is conducted only once.
Model

\[ y_{ij} = \mu + \alpha_j + \beta_i + e_{ij} \]

- \( y_{ij} \) = Observation with A at level j and B at level i
- \( \mu \) = mean response
- \( \alpha_j \) = effect of factor A at level j
- \( \beta_i \) = effect of factor B at level i
- \( e_{ij} \) = error term
Computation of Effects

- Averaging the jth column produces:
  \[ \bar{y}_{.j} = \mu + \alpha_j + \frac{1}{b} \sum_i \beta_i + \frac{1}{b} \sum_i e_{ij} \]

- Since the last two terms are zero, we have:
  \[ \bar{y}_{.j} = \mu + \alpha_j \]

- Similarly, averaging along rows produces:
  \[ \bar{y}_{i.} = \mu + \beta_i \]

- Averaging all observations produces
  \[ \bar{y}_{..} = \mu \]

- Model parameters estimates are:
  \[ \mu = \bar{y}_{..} \]
  \[ \alpha_j = \bar{y}_{.j} - \bar{y}_{..} \]
  \[ \beta_i = \bar{y}_{i.} - \bar{y}_{..} \]

- Easily computed using a tabular arrangement.
## Example 21.1: Cache Comparison

<table>
<thead>
<tr>
<th>Workloads</th>
<th>Two Caches</th>
<th>One Cache</th>
<th>No Cache</th>
</tr>
</thead>
<tbody>
<tr>
<td>ASM</td>
<td>54.0</td>
<td>55.0</td>
<td>106.0</td>
</tr>
<tr>
<td>TECO</td>
<td>60.0</td>
<td>60.0</td>
<td>123.0</td>
</tr>
<tr>
<td>SIEVE</td>
<td>43.0</td>
<td>43.0</td>
<td>120.0</td>
</tr>
<tr>
<td>DHRYSTONE</td>
<td>49.0</td>
<td>52.0</td>
<td>111.0</td>
</tr>
<tr>
<td>SORT</td>
<td>49.0</td>
<td>50.0</td>
<td>108.0</td>
</tr>
</tbody>
</table>
Example 21.1: Computation of Effects

<table>
<thead>
<tr>
<th>Workloads</th>
<th>Two Caches</th>
<th>One Cache</th>
<th>No Cache</th>
<th>Row Sum</th>
<th>Row Mean</th>
<th>Row Effect</th>
</tr>
</thead>
<tbody>
<tr>
<td>ASM</td>
<td>54.0</td>
<td>55.0</td>
<td>106.0</td>
<td>215.0</td>
<td>71.7</td>
<td>-0.5</td>
</tr>
<tr>
<td>TECO</td>
<td>60.0</td>
<td>60.0</td>
<td>123.0</td>
<td>243.0</td>
<td>81.0</td>
<td>8.8</td>
</tr>
<tr>
<td>SIEVE</td>
<td>43.0</td>
<td>43.0</td>
<td>120.0</td>
<td>206.0</td>
<td>68.7</td>
<td>-3.5</td>
</tr>
<tr>
<td>DHRYSTONE</td>
<td>49.0</td>
<td>52.0</td>
<td>111.0</td>
<td>212.0</td>
<td>70.7</td>
<td>-1.5</td>
</tr>
<tr>
<td>SORT</td>
<td>49.0</td>
<td>50.0</td>
<td>108.0</td>
<td>207.0</td>
<td>69.0</td>
<td>-3.2</td>
</tr>
<tr>
<td>Column Sum</td>
<td>255.0</td>
<td>260.0</td>
<td>568.0</td>
<td>1083.0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Column Mean</td>
<td>51.0</td>
<td>52.0</td>
<td>113.6</td>
<td></td>
<td>72.2</td>
<td></td>
</tr>
<tr>
<td>Column effect</td>
<td>-21.2</td>
<td>-20.2</td>
<td>41.4</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

- An average workload on an average processor requires 72.2 ms of processor time.
- The time with two caches is 21.2 ms lower than that on an average processor.
- The time with one cache is 20.2 ms lower than that on an average processor.
- The time without a cache is 41.4 ms higher than the average.
Example 21.1 (Cont)

- Two-cache - One-cache = 1 ms.
- One-cache - No-cache = 41.4-20.2 or 21.2 ms.
- The workloads also affect the processor time required.
- The ASM workload takes 0.5 ms less than the average.
- TECO takes 8.8 ms higher than the average.
**Estimating Experimental Errors**

- **Estimated response:**
  \[ \hat{y}_{ij} = \mu + \alpha_j + \beta_i \]

- **Experimental error:**
  \[ e_{ij} = y_{ij} - \hat{y}_{ij} = y_{ij} - \mu - \alpha_j - \beta_i \]

- **Sum of squared errors (SSE):**
  \[ \text{SSE} = \sum_{i=1}^{b} \sum_{j=1}^{a} e_{ij}^2 \]

- **Example:** The estimated processor time is:
  \[ \hat{y}_{11} = \mu + \alpha_1 + \beta_1 = 72.2 - 21.2 - 0.5 = 50.5 \]

- **Error = Measured-Estimated = 54-50.5 = 3.5**
### Example 21.2: Error Computation

The sum of squared errors is:

\[
SSE = (3.5)^2 + (0.2)^2 + \cdots + (-2.4)^2 = 2368.00
\]

<table>
<thead>
<tr>
<th>Workloads</th>
<th>Two Caches</th>
<th>One Cache</th>
<th>No Cache</th>
</tr>
</thead>
<tbody>
<tr>
<td>ASM</td>
<td>3.5</td>
<td>3.5</td>
<td>-7.1</td>
</tr>
<tr>
<td>TECO</td>
<td>0.2</td>
<td>-0.8</td>
<td>0.6</td>
</tr>
<tr>
<td>SIEVE</td>
<td>-4.5</td>
<td>-5.5</td>
<td>9.9</td>
</tr>
<tr>
<td>DHRystone</td>
<td>-0.5</td>
<td>1.5</td>
<td>-1.1</td>
</tr>
<tr>
<td>SORT</td>
<td>1.2</td>
<td>1.2</td>
<td>-2.4</td>
</tr>
</tbody>
</table>
Example 21.2: Allocation of Variation

\[ y_{ij} = \mu + \alpha_j + \beta_i + e_{ij} \]

- Squaring the model equation:

\[
s_{ij} y_{ij}^2 = ab\mu^2 + b \sum_j \alpha_j^2 + a \sum_i \beta_i^2 + \sum_{ij} e_{ij}^2
\]

\[
SSY = SS0 + SSA + SSB + SSE
\]

\[
SST = SSY - SS0 = SSA + SSB + SSE
\]

13402.41 = 91595 − 78192.59 = 12857.20 + 308.40 + 236.80
100% = 95.9% + 2.3% + 1.8%

- High percent variation explained
  \( \Rightarrow \) Cache choice important in processor design.
Analysis of Variance

- Degrees of freedoms:
  \[ \text{SSY} = \text{SS0} + \text{SSA} + \text{SSB} + \text{SSE} \]
  \[ ab = 1 + (a-1) + (b-1) + (a-1)(b-1) \]

- Mean squares:
  \[ \text{MSA} = \frac{\text{SSA}}{a-1} \]
  \[ \text{MSB} = \frac{\text{SSB}}{b-1} \]
  \[ \text{MSE} = \frac{\text{SSE}}{(a-1)(b-1)} \]
  \[ \frac{\text{MSA}}{\text{MSE}} \sim F[a-1,(a-1)(b-1)] \]

- Computed ratio > \( F_{1-\alpha; a-1,(a-1)(b-1)} \) \( \Rightarrow \) A is significant at level \( \alpha \).
# ANOVA Table

<table>
<thead>
<tr>
<th>Component</th>
<th>Sum of Squares</th>
<th>%Variation</th>
<th>DF</th>
<th>Mean Square</th>
<th>$F$-Comp.</th>
<th>$F$-Table</th>
</tr>
</thead>
<tbody>
<tr>
<td>$y$</td>
<td>$SSY = \sum y_{ij}^2$</td>
<td></td>
<td>$ab$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\bar{y}$</td>
<td>$SS0 = ab\mu^2$</td>
<td></td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$y - \bar{y}$</td>
<td>$SST=SSY-SS0$</td>
<td>100</td>
<td>$ab-1$</td>
<td>$\frac{SSA}{SST}$</td>
<td>$a-1$</td>
<td>$F_{\frac{1-a, a-1}{(a-1)(b-1)}}$</td>
</tr>
<tr>
<td>$A$</td>
<td>$SSA = b\Sigma\alpha_j^2$</td>
<td>100</td>
<td>$a-1$</td>
<td>$\frac{SSA}{a-1}$</td>
<td>$MSA$</td>
<td>$F_{\frac{1-a, a-1}{(a-1)(b-1)}}$</td>
</tr>
<tr>
<td>$B$</td>
<td>$SSB = a\Sigma\beta_i^2$</td>
<td>100</td>
<td>$b-1$</td>
<td>$\frac{SSB}{b-1}$</td>
<td>$MSB$</td>
<td>$F_{\frac{1-a, b-1}{(a-1)(b-1)}}$</td>
</tr>
<tr>
<td>$e$</td>
<td>$SSE = SST-(SSA + SSB)$</td>
<td>$\frac{(a-1)}{b-1}$</td>
<td></td>
<td>$\frac{(a-1)(b-1)}{SSE}$</td>
<td>MSE</td>
<td></td>
</tr>
</tbody>
</table>

Washington University in St. Louis  
CSE567M  
©2011 Raj Jain
### Example 21.3: Cache Comparison

<table>
<thead>
<tr>
<th>Component</th>
<th>Sum of Squares</th>
<th>%Variation</th>
<th>DF</th>
<th>Mean Square</th>
<th>F-Comp.</th>
<th>F-Table</th>
</tr>
</thead>
<tbody>
<tr>
<td>y</td>
<td>91595.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>y..</td>
<td>78192.59</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>y-y..</td>
<td>13402.41</td>
<td>100.0%</td>
<td>14</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Caches</td>
<td>12857.20</td>
<td>95.9%</td>
<td>2</td>
<td>6428.60</td>
<td>217.2</td>
<td>3.1</td>
</tr>
<tr>
<td>Workloads</td>
<td>308.40</td>
<td>2.3%</td>
<td>4</td>
<td>77.10</td>
<td>2.6</td>
<td>2.8</td>
</tr>
<tr>
<td>Errors</td>
<td>236.80</td>
<td>1.8%</td>
<td>8</td>
<td>29.60</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

\[
s_e = \sqrt{\text{MSE}} = \sqrt{29.60} = 5.44
\]

- Cache choice significant.
- Workloads insignificant
Example 21.4: Visual Tests
Confidence Intervals For Effects

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Estimate</th>
<th>Variance</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\mu$</td>
<td>$\bar{y}$..</td>
<td>$s^2_e / ab$</td>
</tr>
<tr>
<td>$\alpha_j$</td>
<td>$\bar{y}_j - \bar{y}$..</td>
<td>$s^2_e (a - 1) / ab$</td>
</tr>
<tr>
<td>$\mu + \alpha_j$</td>
<td>$\bar{y}_j$</td>
<td>$s^2_e / b$</td>
</tr>
<tr>
<td>$\beta_i$</td>
<td>$\bar{y}_i - \bar{y}$..</td>
<td>$s^2_e (b - 1) / ab$</td>
</tr>
<tr>
<td>$\mu + \alpha_j + \beta_i$</td>
<td>$\bar{y}_j + \bar{y}_i - \bar{y}$..</td>
<td>$s^2_e (a + b - 1) / (ab)$</td>
</tr>
<tr>
<td>$\sum_{j=1}^{a} h_j \alpha_j, \sum_{j=1}^{a} h_j = 0$</td>
<td>$\sum_{j=1}^{a} h_j \bar{y}_j$</td>
<td>$s^2_e \sum_{j=1}^{a} h^2_j / b$</td>
</tr>
<tr>
<td>$\sum_{i=1}^{b} h_i \beta_i, \sum_{i=1}^{b} h_i = 0$</td>
<td>$\sum_{i=1}^{b} h_i \bar{y}_i$</td>
<td>$s^2_e \sum_{i=1}^{b} h^2_i / a$</td>
</tr>
<tr>
<td>$\sum_{j=1}^{a} \sum_{i=1}^{b} e^2_{ij} / {(a - 1)(b - 1)}$</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Degrees of freedom for errors = $(a-1)(b-1)$

- For confidence intervals use $t$ values at $(a-1)(b-1)$ degrees of freedom
Example 21.5: Cache Comparison

- Standard deviation of errors:
  \[ s_e = \sqrt{\text{MSE}} = \sqrt{29.60} = 5.4 \]

- Standard deviation of the grand mean:
  \[ s_\mu = s_e / \sqrt{ab} = 5.4 / \sqrt{15} = 1.4 \]

- Standard deviation of \( \alpha \)'s:
  \[ s_{\alpha_j} = s_e \sqrt{(a - 1)/ab} = 5.4 \sqrt{\frac{2}{15}} = 2.8 \]

- Standard deviation of \( \beta \)'s:
  \[ s_{\beta_i} = s_e \sqrt{(b - 1)/ab} = 5.4 \sqrt{\frac{4}{15}} = 2.0 \]
Example 21.5 (Cont)

- Degrees of freedom for the errors are \((a-1)(b-1)=8\). For 90% confidence interval, \(t_{[0.95;8]} = 1.86\).
- Confidence interval for the grand mean:
  \[ 72.2 \mp 1.86 \times 1.4 = 72.2 \mp 2.6 = (69.6, 74.8) \]

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Mean Effect</th>
<th>Std. Dev.</th>
<th>Confidence Interval</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \mu )</td>
<td>72.2</td>
<td>1.4</td>
<td>(69.6, 74.8)</td>
</tr>
<tr>
<td>Caches</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Two Caches</td>
<td>-21.2</td>
<td>2.8</td>
<td>(-24.9, -17.5)</td>
</tr>
<tr>
<td>One Cache</td>
<td>-20.2</td>
<td>2.8</td>
<td>(-23.9, -16.5)</td>
</tr>
<tr>
<td>No Cache</td>
<td>41.4</td>
<td>2.8</td>
<td>(37.7, 45.1)</td>
</tr>
</tbody>
</table>

- All three cache alternatives are significantly different from the average.
All workloads, except TECO, are similar to the average and hence to each other.
Example 21.5: CI for Differences

<table>
<thead>
<tr>
<th></th>
<th>Two Caches</th>
<th>One Cache</th>
<th>No Cache</th>
</tr>
</thead>
<tbody>
<tr>
<td>Two Caches</td>
<td>( -7.4, 5.4)†</td>
<td>( -69.0, -56.2)</td>
<td></td>
</tr>
<tr>
<td>One Cache</td>
<td></td>
<td>( -68.0, -55.2)</td>
<td></td>
</tr>
</tbody>
</table>

† ⇒ Not significant

- Two-cache and one-cache alternatives are both significantly better than a no cache alternative.
- There is no significant difference between two-cache and one-cache alternatives.
Multiplicative Models

- Additive model:
  \[ y_i = \mu + \alpha_j + \beta_i + e_{ij} \]
- If factors multiply ⇒ Use multiplicative model
- Example: processors and workloads
  - Log of response follows an additive model
- If the spread in the residuals increases with the mean response ⇒ Use transformation
Missing Observations

- **Recommended Method:**
  - Divide the sums by respective number of observations
  - Adjust the degrees of freedoms of sums of squares
  - Adjust formulas for standard deviations of effects

- **Other Alternatives:**
  - Replace the missing value by $\hat{y}$ such that the residual for the missing experiment is zero.
  - Use $y$ such that SSE is minimum.
Two Factor Designs Without Replications

- Model:
  \[ y_{ij} = \mu + \alpha_j + \beta_i + e_{ij} \]

- Effects are computed so that:
  \[ \sum_{j=1}^{a} \alpha_j = 0 \]
  \[ \sum_{i=1}^{b} \beta_i = 0 \]

- Effects:
  \[ \mu = \bar{y}_{..} ; \quad \alpha_j = \bar{y}_{.j} - \bar{y}_{..} ; \quad \beta_i = \bar{y}_{i.} - \bar{y}_{..} \]
Summary (Cont)

- Allocation of variation: SSE can be calculated after computing
  \[ \sum_{i,j} y_{i,j}^2 = ab\mu^2 + b \sum_j \alpha_j^2 + a \sum_i \beta_i^2 + \sum_{i,j,k} e_{i,j,k}^2 \]
  \[ \text{SSY} = \text{SS0} + \text{SSA} + \text{SSB} + \text{SSE} \]

  Degrees of freedom:
  \[ \text{SSY} = \text{SS0} + \text{SSA} + \text{SSB} + \text{SSE} \]
  \[ ab = 1 + (a - 1) + (b - 1) + (a - 1)(b - 1) \]

- Mean squares:
  \[ \text{MSA} = \frac{\text{SSA}}{a - 1}; \text{MSB} = \frac{\text{SSB}}{b - 1}; \text{MSE} = \frac{\text{SSE}}{(a - 1)(b - 1)} \]

- Analysis of variance:
  \[ \text{MSA/MSE should be greater than } F_{[1-\alpha; a-1,(a-1)(b-1)]}. \]
  \[ \text{MSB/MSE should be greater than } F_{[1-\alpha; b-1,(a-1)(b-1)]}. \]
Summary (Cont)

- Standard deviation of effects:
  \[ s^2_\mu = \frac{s^2_e}{ab}; \quad s^2_{\alpha_j} = \frac{s^2_e (a - 1)}{ab}; \quad s^2_{\beta_\ell} = \frac{s^2_e (b - 1)}{ab}; \]

- Contrasts:
  For \( \sum_{j=1}^{a} h_j \alpha_j, \sum_{j=1}^{a} h_j = 0 \):
    - Mean = \( \sum_{j=1}^{a} h_j \bar{y}_j \); Variance = \( s^2_e \sum_{j=1}^{a} h_j^2 / b \)
  For \( \sum_{i=1}^{b} h_i \beta_i, \sum_{i=1}^{b} h_i = 0 \):
    - Mean = \( \sum_{i=1}^{b} h_i \bar{y}_i \); Variance = \( s^2_e \sum_{i=1}^{b} h_i^2 / a \)

- All confidence intervals are calculated using \( t_{[1-\alpha/2;(a-1)(b-1)]} \).

- Model assumptions:
  - Errors are IID normal variates with zero mean.
  - Errors have the same variance for all factor levels.
  - The effects of various factors and errors are additive.

- Visual tests:
  - No trend in scatter plot of errors versus predicted responses
  - The normal quantile-quantile plot of errors should be linear.
Homework 21: Exercise 21.1

Execution Times

<table>
<thead>
<tr>
<th>Workloads</th>
<th>Scheme86</th>
<th>Spectrum125</th>
<th>Spectrum62.5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Garbage Collection</td>
<td>39.97</td>
<td>99.06</td>
<td>56.24</td>
</tr>
<tr>
<td>Pattern Match</td>
<td>0.958</td>
<td>1.672</td>
<td>1.252</td>
</tr>
<tr>
<td>Bignum Addition</td>
<td>0.01910</td>
<td>0.03175</td>
<td>0.01844</td>
</tr>
<tr>
<td>Bignum Multiplication</td>
<td>0.256</td>
<td>0.423</td>
<td>0.236</td>
</tr>
<tr>
<td>Fast Fourier Transform (1024)</td>
<td>10.21</td>
<td>20.28</td>
<td>10.14</td>
</tr>
</tbody>
</table>

Analyze the data of Case study 21.2 using a 2-factor additive model.

- Estimate effects and prepare ANOVA table
- Plot residuals as a function of predicted response.
- Also, plot a normal quantile-quantile plot for the residuals.
- Determine 90% confidence intervals for the paired differences.
  (Confidence intervals of $\alpha_1-\alpha_2$, $\alpha_1-\alpha_3$, $\alpha_2-\alpha_3$)
- Are the processors significantly different?
- Discuss what indicators in the data, analysis, or plot would suggest that this is not a good model.