

EEC 686/785  
Modeling & Performance Evaluation of  
Computer Systems

## Lecture 12

Wenbing Zhao

Department of Electrical and Computer Engineering  
Cleveland State University

wenbing@ieee.org

(based on Dr. Raj Jain's lecture notes)



## Outline

- Review of lecture 11
- $2^k r$  Factorial Designs with Replication



## Terminology

- **Response variables:** outcome
- **Factors:** variables that affect the response variable
- **Levels:** the values that a factor can assume
- **Primary factors** and **secondary factors**
- **Replication:** repetition of all or some experiments
- **Design:** the number of experiments, the factor level and number of replications for each experiment
- **Experimental Unit**
- **Interaction:** effect of one factor depends upon the level of the other



## Types of Experimental Designs

- **Simple Designs:** vary one factor at a time
- **Full Factorial Design:** all combinations
- **Fractional Factorial Designs:** use only a fraction of the full factorial design

## 2<sup>k</sup> 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 of a factor is unidirectional, i.e., the performance either continuously decreases or continuously increases as the factor is increased from min to max
  - E.g., memory size, the number of disk drives

## 2<sup>2</sup> Factorial Designs: Model

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

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

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

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

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

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

- Unique solution for  $q_A$  and  $q_B$ :  
 $y = 40 + 20x_A + 10x_B + 5x_A x_B$

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

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

- Interpretation:
  - Mean performance = 40 MIPS
  - Effect of memory = 20 MIPS
  - Effect of cache = 10 MIPS
  - Interaction between memory and cache = 5 MIPS

## Allocation of Variation

- Importance of a factor = proportion of the *variation* explained
- Sample Variance of  $y = s_y^2 = \frac{\sum_{i=1}^{2^2} (y_i - \bar{y})^2}{2^2 - 1}$
- Variation of  $y \triangleq$  Numerator  

$$= \sum_{i=1}^{2^2} (y_i - \bar{y})^2$$

$$= \text{sum of squares total (SST)}$$
- For a 2<sup>2</sup> design  $SST = 2^2 q_A^2 + 2^2 q_B^2 + 2^2 q_{AB}^2$
- Variation due to A=SSA=  $2^2 q_A^2$ , etc.

## Derivation

- Model:

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

- Notice

- The sum of entries in each column is zero
- The sum of the squares of entries in each column is 4
- **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

- Variation of  $y$

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

## $2^k r$ Factorial Designs with Replication

- $r$  replications of  $2^k$  experiments
  - $> 2^k r$  observations
  - $>$  Allows estimation of experimental errors
- Model:  $y = q_0 + q_A x_A + q_B x_B + q_{AB} x_A x_B + e$ 
  - $> e = \text{experimental error}$

## Computation of Effects

- Memory-Cache Study: Simply use means of  $r$  measurements

$I$	$A$	$B$	$AB$	$y$	Mean $\bar{y}$
1	-1	-1	1	(15, 18, 12)	15
1	1	-1	-1	(45, 48, 51)	48
1	-1	1	-1	(25, 28, 19)	24
1	1	1	1	(75, 75, 81)	77
164	86	38	20		total
41	21.5	9.5	5		total/4

- $>$  Effects:  $q_0 = 41, q_A = 21.5, q_B = 9.5, q_{AB} = 5$

## Estimation of Experimental Errors

- Estimated Response:
 
$$\hat{y}_i = q_0 + q_A x_{Ai} + q_B x_{Bi} + q_{AB} x_{Ai} x_{Bi}$$
- Experimental error = estimated - measured
 
$$e_{ij} = y_{ij} - \hat{y}_i$$

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

$$\sum_{i,j} e_{ij} = 0$$
- Sum of squared errors:  $SSE = \sum_{i=1}^{2^2 r} e_i^2$

## Experimental Errors: Example

- Estimated Response:

$$\hat{y}_1 = q_0 - q_A - q_B + q_{AB} = 41 - 21.5 - 9.5 + 5 = 15$$

- Experimental errors:

$$e_{11} = y_{11} - \hat{y}_1 = 15 - 15 = 0$$

i	Effect				Estimated Response				Measured Responses			Errors		
	T	A	B	AB	$\hat{y}_i$	$y_{i1}$	$y_{i2}$	$y_{i3}$	$e_{i1}$	$e_{i2}$	$e_{i3}$			
	41	21.5	9.5	5	$\hat{y}_i$	$y_{i1}$	$y_{i2}$	$y_{i3}$	$e_{i1}$	$e_{i2}$	$e_{i3}$			
1	1	-1	-1	1	15	15	18	12	0	3	-3			
2	1	1	-1	-1	48	45	48	51	-3	0	3			
3	1	-1	1	-1	24	25	28	19	1	4	-5			
4	1	1	1	1	77	75	75	81	-2	-2	4			

$$SSE = 0^2 + 3^2 + (-3)^2 + (-3)^2 + \dots + 4^2 = 102$$

## Allocation of Variation

- $\bar{y}_{..}$  denotes the mean of responses from all replications of all experiments
  - The dots in the subscript indicate the dimension along which the averaging is done
  - $\bar{y}_{.i}$  denotes the means of responses in all replications of the  $i$ th experiment
- Total variation or **total sum of squares (SST)**:

$$SST = \sum_{ij} (y_{ij} - \bar{y}_{..})^2$$

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

$$\sum_{ij} (y_{ij} - \bar{y}_{..})^2 = 2^2 r q_A^2 + 2^2 r q_B^2 + 2^2 r q_{AB}^2 + \sum_{i,j} e_{ij}^2$$

$$SST = SSA + SSB + SSAB + SSE$$

## Derivation

- Model:

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

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

- Since  $x$ 's, their products, and all errors add to 0

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

- Mean response:

$$\bar{y}_{..} = \frac{1}{2^2 r} \sum_{i,j} y_{ij} = q_0$$

- Squaring both sides of the model and ignoring cross product terms:

$$\sum_{i,j} y_{ij}^2 = \sum_{i,j} q_0^2 + \sum_{i,j} q_A^2 x_{Ai}^2 + \sum_{i,j} q_B^2 x_{Bi}^2 + \sum_{i,j} q_{AB}^2 x_{Ai}^2 x_{Bi}^2 + \sum_{i,j} e_{ij}^2$$

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

## Derivation

- Total variation:

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

$$= \sum_{i,j} y_{ij}^2 - \sum_{i,j} y_{ij} \bar{y}_{..}$$

$$= SSY - SS0$$

$$= SSA + SSB + SSAB + SSE$$

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

## Example: Memory-Cache Study

$$SSY = 15^2 + 18^2 + 12^2 + 45^2 + \dots + 75^2 + 75^2 + 81^2 = 27204$$

$$SS0 = 2^2 r q_0^2 = 12 \times 41^2 = 20172$$

$$SSA = 2^2 r q_A^2 = 12 \times (21.5)^2 = 5547$$

$$SSB = 2^2 r q_B^2 = 12 \times (9.5)^2 = 1083$$

$$SSAB = 2^2 r q_{AB}^2 = 12 \times 5^2 = 300$$

$$SSE = 27204 - 2^2 \times 3(41^2 + 21.5^2 + 9.5^2 + 5^2) = 102$$

$$SST = SSY - SS0 = 27204 - 20172 = 7032$$

- $SSA + SSB + SSAB + SSE = 5547 + 1083 + 300 + 102 = 7032 = SST$

## Example: Memory-Cache Study

- Factor A explains 5547/7032 or 78.88%
- Factor B explains 15.40%
- Interaction AB explains 4.27%
- 1.45% is unexplained and is attributed to errors

## Confidence Intervals for Effects

- Effects are random variables
- Errors  $\sim N(0, \sigma_e) \Rightarrow y \sim N(\bar{y}_{..}, \sigma_e)$

$$q_0 = \frac{1}{2^2 r} \sum_{i,j} y_{ij}$$

- $q_0 =$  Linear combination of normal variates  $\Rightarrow$   
 $q_0$  is normal with variance  $\sigma_e^2 / (2^2 r)$

## Confidence Intervals for Effects

- Variance of errors:

$$s_e^2 = \frac{1}{2^2(r-1)} \sum_{ij} e_e^2 = \frac{SSE}{2^2(r-1)} \triangleq MSE$$

- **Mean Square of Errors (MSE):** quantity at the right hand side
- Denominator =  $2^2(r-1)$  = # of independent terms in SSE  $\Rightarrow$  degrees of freedom SSE has

## Confidence Intervals for Effects

- Estimated variance of  $q_0$ :  $s_{q_0}^2 = s_e^2 / (2^2 r)$
- Similarly,  $s_{q_A} = s_{q_B} = s_{q_{AB}} = s_e / \sqrt{2^2 r}$
- Confidence intervals (CI) for the effects:
 
$$q_i \mp t_{[1-\alpha/2; 2^2(r-1)]} s_{q_i}$$
- CI does not include a zero => significant

## Example

- For 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:

$$s_{q_i} = \frac{s_e}{\sqrt{2^2 r}} = \frac{3.57}{\sqrt{12}} = 1.03$$

- For 90% confidence:  $t_{[0.95, 8]} = 1.86$

## Example

- Confidence intervals:
  - >  $q_i \pm (1.86)(1.03) = q_i \pm 1.92$
  - >  $q_0 = (39.08, 42.91)$
  - >  $q_A = (19.58, 23.41)$
  - >  $q_B = (7.58, 11.41)$
  - >  $q_{AB} = (3.08, 6.91)$
  - > No zero crossing => all effects are significant

## Confidence Intervals for Contrasts

- It is also possible to compute variance and confidence intervals for any contrast of effects.
- A **contrast** is any linear combination whose coefficients add up to zero
- Variance of  $\sum h_i q_i$ , where  $h_i=0$ , is:

$$s_{\sum h_i q_i}^2 = \frac{s_e^2 \sum h_i^2}{2^2 r}$$

- For 100(1- $\alpha$ )% confidence interval, use  $t_{[1-\alpha/2; 2^2(r-1)]}$

## Example 18.5

- Memory-cache study: the confidence interval for  $u = q_A + q_B - 2q_{AB}$  is calculated as follows:
  - Coefficients = 0, 1, 1, and -2  $\Rightarrow$  Contrast
  - Mean  $\bar{u} = 21.5 + 9.5 - 2 \times 5 = 11$
  - Variance  $s_u^2 = \frac{s_e^2 \times 6}{2^2 \times 3} = 6.375$
  - Standard deviation  $s_u = \sqrt{6.375} = 2.52$
  - $t_{[0.95; 8]} = 1.86$
  - 90% confidence interval for u:
 
$$\bar{u} \mp t_{s_u} = 11 \mp 1.86 \times 2.52 = (6.31, 15.69)$$

## CI for Predicted Responses

- Mean response  $\hat{y} = q_0 + q_A x_A + q_B x_B + q_{AB} x_A x_B$
- The standard deviation of the mean of  $m$  responses:

$$s_{\hat{y}_m} = s_e \left( \frac{1}{n_{eff}} + \frac{1}{m} \right)^{1/2}$$

$n_{eff}$  = effective degree of freedom

$$= \frac{\text{total number of runs}}{1 + \text{sum of DFs of params used in } \hat{y}}$$

$$= \frac{2^2 r}{5}$$

## CI for Predicted Responses

- 100(1- $\alpha$ )% confidence interval:

$$\hat{y} \mp t_{[1-\alpha/2; 2^2(r-1)]} s_{\hat{y}_m}$$

- A single run ( $m=1$ ):  $s_{\hat{y}_1} = s_e \left( \frac{5}{2^2 r} + 1 \right)^{1/2}$
- Population mean ( $m=\infty$ ):  $s_{\hat{y}} = s_e \left( \frac{5}{2^2 r} \right)^{1/2}$

## Example: Memory-Cache Study

- For  $x_A = -1$  and  $x_B = -1$ :
  - A single confirmation experiment:
 
$$\hat{y}_1 = q_0 - q_A - q_B + q_{AB} = 41 - 21.5 - 9.5 + 5 = 15$$
 Standard deviation of the predication:
 
$$s_{\hat{y}_1} = s_e \left( \frac{5}{2^2 r} + 1 \right)^{1/2} = 3.57 \sqrt{\frac{5}{12} + 1} = 4.25$$
  - Using  $t_{[0.95; 8]} = 1.86$ , the 90% confidence interval is:
 
$$15 \pm 1.86 \times 4.25 = (8.09, 22.91)$$

## Example: Memory-Cache Study

- For  $x_A = -1$  and  $x_B = -1$  (continued):

- **Mean response for 5 experiments in future:**

$$s_{\hat{y}_i} = s_e \left( \frac{5}{2^2 r} + \frac{1}{m} \right)^{1/2} = 3.57 \sqrt{\frac{5}{12} + \frac{1}{5}} = 2.80$$

The 90% confidence interval is:

$$15 \pm 1.86 \times 2.80 = (9.79, 20.29)$$

## Example: Memory-Cache Study

- For  $x_A = -1$  and  $x_B = -1$  (continued):

- **Mean response for a large number of experiments in future:**

$$s_{\hat{y}} = s_e \left( \frac{5}{2^2 r} \right)^{1/2} = 3.57 \sqrt{\frac{5}{12}} = 2.30$$

- The 90% confidence interval is:

$$15 \pm 1.86 \times 2.30 = (10.72, 19.28)$$

## Example: Memory-Cache Study

- For  $x_A = -1$  and  $x_B = -1$  (continued):

- **Current mean response:** not for future (use the formula for contrasts):

$$s_{\hat{y}} = \sqrt{\frac{s_e \sum h_i^2}{2^2 r}} = \sqrt{\frac{12.75 \times 4}{12}} = 2.06$$

The 90% confidence interval is:

$$15 \pm 1.86 \times 2.06 = (11.17, 18.83)$$

- Notice: Confidence intervals become narrower

## Visual Tests for Verifying the Assumptions

- In deriving the expressions for effects, we made essentially the same assumptions as in regression analysis:

- Errors are statistically independent
- Errors are additive
- Errors are normally distributed
- Errors have a constant standard deviation
- Effects of factors are additive

⇒ Observations are independent and normally distributed with constant variance

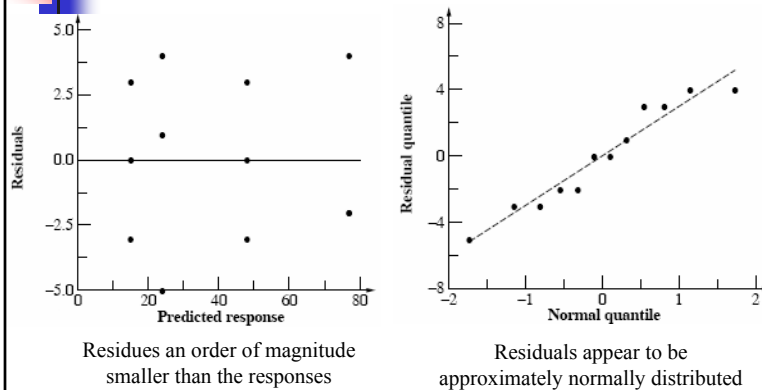
## Visual Tests for Verifying the Assumptions

- Visual tests for independent errors:
  - Scatter plot of residuals versus the predicted response
  - Magnitude of residuals < magnitude of responses/10 => ignore trends
  - Plot the residuals as a function of the experiment number
  - Trend up or down => other factors or side effects
- Visual tests for Normally distributed errors:
  - Prepare normal quantile-quantile plot of errors
  - If the plot is approximately linear, the assumption is satisfied

## Visual Tests for Verifying the Assumptions

- Visual tests for constant standard deviation of errors:
  - Scatter plot of  $y$  for various levels of the factor
  - Spread at one level significantly different than that at other => Assumption of constant variance is not valid, need transformation

## Visual Tests Example: Memory-Cache



## Multiplicative Models for $2^2r$ Experiments

- **Additive model:**

$$y_{ij} = q_0 + q_A x_A + q_B x_B + q_{AB} x_A x_B + e_{ij}$$
  - Not valid if effects do not add
- E.g., **execution time of workloads**
  - $i$ th processor speed =  $v_i$  instructions/second
  - $j$ th workload size =  $w_j$  instructions
  - Execution time  $y_{ij} = v_i \times w_j$
  - The two effects **multiply**. Logarithm => additive model:
 
$$\log(y_{ij}) = \log(v_i) + \log(w_j)$$

## Multiplicative Models for $2^2r$ Experiments

- **Correct model:**  $y'_{ij} = q_0 + q_A x_A + q_B x_B + q_{AB} x_A x_B + e_{ij}$

where,  $y'_{ij} = \log(y_{ij})$

- **Taking an antilog of additive effects  $q$  to get the multiplicative effects  $u$ :**

$$u_A = 10^{q_A}, u_B = 10^{q_B}, \text{ and } u_{AB} = 10^{q_{AB}}$$

- $u_A$  = ratio of MIPS rating of the two processors
- $u_B$  = ratio of the size of the two workload
- Antilog of additive mean  $q_0 \Rightarrow$  geometric mean

$$\dot{y} = 10^{q_0} = (y_1 y_2 \cdots y_n)^{1/n} \quad n = 2^2 r$$

## Example: Execution Times

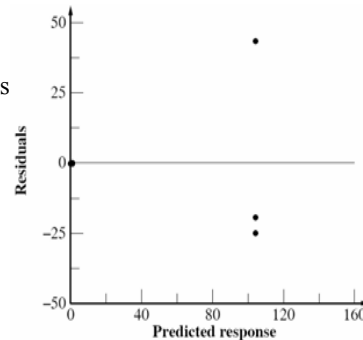
Analysis Using an Additive Model						
$I$	$A$	$B$	$AB$	$y$	Mean $\bar{y}$	
1	-1	-1	1	(85.10, 79.50, 147.90)	104.170	
1	1	-1	-1	(0.891, 1.047, 1.072)	1.003	
1	-1	1	-1	(0.955, 0.933, 1.122)	1.003	
1	1	1	1	(0.0148, 0.0126, 0.0118)	0.013	
				total	106.19	-104.15
				total/4	26.55	-26.04
						102.17
						25.54

- Additive model is not valid because of physical consideration
  - Effects of workload and processors do not add. They multiply
- Additive model is not valid because of large range for  $y$ 
  - $y_{max}/y_{min} = 147.90/0.0118$  or  $12,534 \Rightarrow$  log transformation
  - Taking an arithmetic mean of 114.17 and 0.013 is inappropriate

## Example: Execution Times

- Additive model is not valid because:

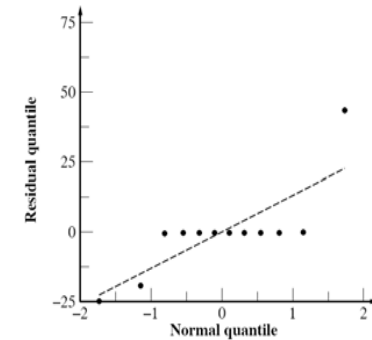
- The residuals are not small as compared to the response
- The spread of residuals is large at larger value of the response  $\Rightarrow$  log transformation



## Example: Execution Times

- Additive model is not valid because:

- The residual distribution has a longer tail than normal



## Analysis Using Multiplicative Model

- Transformed data for multiplicative model example

Data After Log Transformation					
<i>I</i>	<i>A</i>	<i>B</i>	<i>AB</i>	<i>y</i>	Mean $\bar{y}$
1	-1	-1	1	( 1.93, 1.90, 2.17)	2.00
1	1	-1	-1	( -0.05, 0.02, 0.03)	0.00
1	-1	1	-1	( -0.02, -0.03, 0.05)	0.00
1	1	1	1	( -1.83, -1.90, -1.93)	-1.89
0.11	-3.89	-3.89	0.11		total
0.03	-0.97	-0.97	0.03		total/4

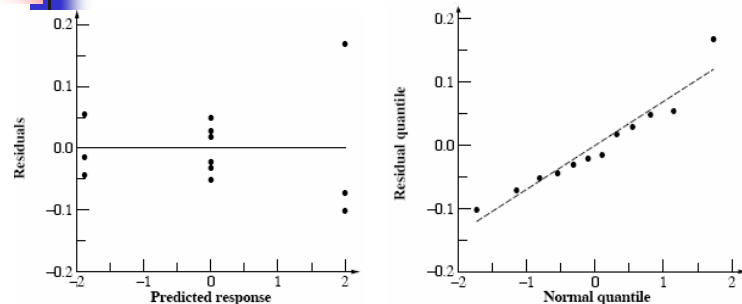
## Analysis Using Multiplicative Model

Factor	Additive Model			Multiplicative Model		
	Effect	% Var.	Conf. Interval	Effect	% Var.	Conf. Interval
<i>I</i>	26.55		( 16.35, 36.74)	0.03		( -0.02, 0.07)
<i>A</i>	-26.04	30.1%	( -36.23, -15.84)	-0.97	49.9%	( -1.02, -0.93)
<i>B</i>	-26.04	30.1%	( -36.23, -15.84)	-0.97	49.9%	( -1.02, -0.93)
<i>AB</i>	25.54	29.0%	( 15.35, 35.74)	0.03	0.0%	( -0.02, 0.07)
<i>e</i>		10.8%			0.2%	

† ⇒ Not Significant

- Percentage of variation explained by the two models
- With multiplicative model
  - Interaction is almost zero
  - Unexplained variation is only 0.2%

## Using Multiplicative: Visual Tests



- Conclusion: multiplicative model is better than the additive model

## Using Multiplicative: Interpretation of Results

$$\begin{aligned} \log(y) &= q_0 + q_A x_A + q_B x_B + q_{AB} x_A x_B + e \\ y &= 10^{q_0} 10^{q_A x_A} 10^{q_B x_B} 10^{q_{AB} x_A x_B} 10^e \\ &= 10^{0.03} 10^{-0.97 x_A} 10^{-0.97 x_B} 10^{0.03 x_A x_B} 10^e \\ &= 1.07 \times 0.107^{x_A} \times 0.107^{x_B} \times 1.07^{x_A x_B} 10^e \end{aligned}$$

- The time for an average processor on an average benchmark is 1.07
- The time on processor  $A_1$  is 9 times ( $0.107^{-1}$ ) that on an average processor. The time on  $A_2$  is one ninth ( $0.107^1$ ) of that on an average processor

## Using Multiplicative: Interpretation of Results

- MIPS rate for  $A_2$  is 81 times that of  $A_1$
- Benchmark  $B_1$  executes 81 times more instructions than  $B_2$
- The interaction is negligible  
=> Results apply to all benchmarks and processors

## Transformation Considerations

- $y_{max}/y_{min}$  small => multiplicative model results similar to additive model
- Many other transformation possible

## Transformation Considerations

### Box-Cox family of transformations:

$$w = \begin{cases} \frac{y^a - 1}{ag^{a-1}}, & a \neq 0 \\ (\ln y)g, & a = 0 \end{cases}$$

Where  $g$  is the geometric mean of the responses:

$$g = (y_1 y_2 \cdots y_n)^{1/n}$$

- $w$  has the same units as  $y$
- $a$  can have any real value, positive, negative, or zero
- Plot SSE as a function of  $a$  => optimal  $a$
- Knowledge about the system behavior should always take precedence over statistical considerations

## General $2^k r$ Factorial Design

- Model:  $y_{ij} = q_0 + q_A x_A + q_B x_B + q_{AB} x_A x_B + \cdots + e_{ij}$

- Parameter estimation:  $q_j = \frac{1}{2^k} \sum_{i=1}^{2^k} S_{ij} \bar{y}_i$

$S_{ij} = (i, j)$ th entry in the sign table

- Sum of squares:  $SSY = \sum_{i=1}^{2^k} \sum_{j=1}^r y_{ij}^2$

$$SS0 = 2^k r q_0^2$$

$$SST = SSY - SS0$$

$$SSj = 2^k r q_j^2 - \sum_{j=1}^{2^k-1} SSj$$

## General $2^k r$ Factorial Design

- Percentage of  $y$ 's variation explained by  $j$ th effect  
 $= (SS_j/SST) \times 100\%$

- Standard deviation of errors:  $s_e = \sqrt{\frac{SSE}{2^k(r-1)}}$

- Standard deviation of effects:

$$s_{q_0} = s_{q_A} = s_{q_B} = s_{q_{AB}} = s_e / \sqrt{2^k r}$$

- Variance of contrast  $\sum h_i q_i$ , where  $h_i=0$ , is:  $s_{\sum h_i q_i}^2 = \frac{s_e^2 \sum h_i^2}{2^k r}$

- Standard deviation of the mean of  $m$  future responses:

$$s_{\bar{y}_m} = s_e \left( \frac{1+2^k}{2^k r} + \frac{1}{m} \right)^{1/2}$$

## General $2^k r$ Factorial Design

- Confidence intervals are calculated using  $t_{[1-\alpha/2; 2^k(r-1)]}$

- Modeling assumptions

- Errors are IID (independently and identically distributed) normal variates with zero mean
- Errors have the same variance for all values of the predictors
- Effects and errors are additive

## General $2^k r$ Factorial Design

- Visual tests

- The scatter plot of errors versus predicted responses should not have any trend
- The normal quantile-quantile plot of errors should be linear
- Spread of  $y$  values in all experiments should be comparable
- If any of the above visual tests fail or if the ratio  $y_{\max}/y_{\min}$  is large, a multiplicative model should be investigated

## Example: A $2^3$ Design

$I$	$A$	$B$	$C$	$AB$	$AC$	$BC$	$ABC$	$y$	Mean $\bar{y}$
1	-1	-1	-1	1	1	1	-1	(14, 16, 12)	14
1	1	-1	-1	-1	-1	1	1	(22, 18, 20)	20
1	-1	1	-1	-1	1	-1	-1	(11, 15, 19)	15
1	1	1	-1	1	-1	-1	-1	(34, 30, 35)	33
1	-1	-1	1	1	-1	-1	1	(46, 42, 44)	44
1	1	-1	1	-1	1	-1	-1	(58, 62, 60)	60
1	-1	1	1	-1	-1	1	-1	(50, 55, 54)	53
1	1	1	1	1	1	1	1	(86, 80, 74)	80
319	67	43	155	23	19	15	-1	total	
39.87	8.375	5.375	19.37	2.875	2.375	1.875	-0.125	total/8	

## Example: A 2<sup>3</sup> Design

### ■ Sums of squares

Component	Sum of Squares	Percent Variation
$y$	4.9E4	
$\bar{y}$	3.8E4	
$y - \bar{y}$	1.1E4	100.00%
$A$	1683.0	14.06%
$B$	693.3	5.79%
$C$	9009.0	75.27%
$AB$	198.3	1.66%
$AC$	135.4	1.13%
$BC$	84.4	0.70%
$ABC$	0.4	0.00%
Errors	164.0	1.37%

## Example: A 2<sup>3</sup> Design

### ■ Confidence intervals of effects

- > The errors have 2<sup>3</sup>(3-1) or 16 degrees of freedom. Standard deviation of errors:

$$s_e = \sqrt{\frac{SSE}{2^k(r-1)}} = \sqrt{\frac{164}{16}} = 3.20$$

- > Standard deviation of effects:

$$s_{q_i} = s_e / \sqrt{2^3 3} = 3.20 / \sqrt{24} = 0.654$$

- >  $t_{[0.95,16]} = 1.337$

## Example: A 2<sup>3</sup> Design

- > 90% confidence intervals for parameters:

$$q_i \pm (1.337)(0.654) = q_i \pm 0.874$$

- ✦  $q_0 = (39.00, 40.74)$
- ✦  $q_A = (7.50, 9.25)$
- ✦  $q_B = (4.50, 6.25)$
- ✦  $q_C = (18.50, 20.24)$
- ✦  $q_{AB} = (2.00, 3.75)$
- ✦  $q_{AC} = (1.50, 3.25)$
- ✦  $q_{BC} = (1.00, 2.75)$
- ✦  $q_{ABC} = (-1.00, 0.75)$

All effects except  $q_{ABC}$  are significant

## Example: A 2<sup>3</sup> Design

### ■ Predication

- > For a single confirmation experiment ( $m=1$ )
- > With  $A = B = C = -1$ :

$$\hat{y} = 14$$

$$s_{\hat{y}} = s_e \left( \frac{5}{2^k r} + \frac{1}{m} \right)^{1/2} = 3.2 \left( \frac{5}{24} + 1 \right)^{1/2} = 3.52$$

- > 90% confidence interval:

$$14 \pm 1.337 \times 3.52 = 14 \pm 4.70 = (9.30, 18.70)$$

## Case Study: Garbage Collection

Factors and Levels

Variable	Factor	-1	1
A	Workload	Single Task	Several parallel tasks
B	Compiler	Simple	Deallocating
C	Limbo List	Enabled	Disabled
D	Chunk Size	4K bytes	16K bytes

## Case Study: Garbage Collection

### Measured Data

<i>I</i>	<i>A</i>	<i>B</i>	<i>C</i>	<i>D</i>	<i>y</i>	Mean $\bar{y}$
1	-1	-1	-1	-1	( 97, 97, 97)	97.00
1	1	-1	-1	-1	( 31, 31, 32)	31.33
1	-1	1	-1	-1	( 97, 97, 97)	97.00
1	1	1	-1	-1	( 31, 32, 31)	31.33
1	-1	-1	1	-1	( 97, 97, 97)	97.00
1	1	-1	1	-1	( 32, 32, 31)	31.67
1	-1	1	1	-1	( 97, 97, 97)	97.00
1	1	1	1	-1	( 32, 32, 32)	32.00
1	-1	-1	-1	1	( 407, 407, 407)	407.00
1	1	-1	-1	1	( 135, 136, 135)	135.33
1	-1	1	-1	1	( 409, 409, 409)	409.00
1	1	1	-1	1	( 135, 135, 136)	135.33
1	-1	-1	1	1	( 407, 407, 407)	407.00
1	1	-1	1	1	( 139, 140, 139)	139.33
1	-1	1	1	1	( 409, 409, 409)	409.00
1	1	1	1	1	( 139, 139, 140)	139.33
					2695.67	
					-1344.33	
					4.33	
					9.00	
					1667.00	
					168.48	
					-84.02	
					0.27	
					0.56	
					104.19	
						total
						total/8

## Case Study: Garbage Collection

### Effects and variation explained

Factor	Effect	% Variation	Conf. Interval
<i>I</i>	168.48	138.1%	( 168.386, 168.573)
<i>A</i>	-84.02	34.4%	( -84.114, -83.927)
<i>B</i>	0.27	0.0%	( 0.177, 0.364)
<i>C</i>	0.56	0.0%	( 0.469, 0.656)
<i>D</i>	104.19	52.8%	( 104.094, 104.281)
<i>AB</i>	-0.23	0.0%	( -0.323, -0.136)
<i>AC</i>	0.56	0.0%	( 0.469, 0.656)
<i>AD</i>	-51.31	12.8%	( -51.406, -51.219)
<i>BC</i>	0.02	0.0%	( -0.073, 0.114)†
<i>BD</i>	0.23	0.0%	( 0.136, 0.323)
<i>CD</i>	0.44	0.0%	( 0.344, 0.531)
<i>ABC</i>	0.02	0.0%	( -0.073, 0.114)†
<i>ABD</i>	-0.27	0.0%	( -0.364, -0.177)
<i>ACD</i>	0.44	0.0%	( 0.344, 0.531)
<i>BCD</i>	-0.02	0.0%	( -0.114, 0.073)†
<i>ABCD</i>	-0.02	0.0%	( -0.114, 0.073)†

† ⇒ Not Significant

## Case Study: Garbage Collection

### Conclusions

- Most of the variation is explained by factors A (workload), D (chunk size), and the interaction AD between the two
- The variation due to experimental error is small ⇒ Several effects that explain less than 0.05% of variation (listed as '0.0%') are statistically significant
- Only effects A, D, and AD are **both practically significant and statistically significant**