Computer Systems Performance Analysis and Benchmarking (37-235)

Analytic Modeling

Simulation

Measurements / Benchmarking

Lecture by:
Prof. Thomas Stricker

Assignments/Projects:
Christian Kurmann

Textbook:
Raj Jain, "The Art of Computer Systems Performance Analysis", 1991 Wiley & Sons, New York

Topic of Today:

- **Introduction Experimental Design**
- **2^k Factorial Design**
- **2^kr Factorial Design**

Experimental Design

Goal:

- Find out which factor contributes what before and during the analysis and not just after the analysis.

Example: Workstation - Model 1982

- CPU

  {68000, Z80, 8086}
- DRAM Memory

  [512kB, 2MB, 8MB]
- Disk drives

  [1,2,3,4]
- Workloads

  {secretarial, managerial, scientific}
- Users

  {high-school, college, post-graduate}

Terminology

- Response Variable

  The outcome of an experiment, the performance achieved.
- Factors

  Variables in the Experiment

  Predictor Variables
  - Primary factors
    Need to be quantified
    Are of interest
  - Secondary factors
    Influence performance
    Are not of interest
- Levels

  Possible Values of Factors

  Set or range, discrete or continuous

- Replication

  Number of repetitions of each experiment.
- Design

  Recipe to determine # of experiments

  e.g. full factorial with replication 5
  3x3x4x3x3=324*5=1215
- Experimental unit

  Entity used for experiment

  e.g. different workstations, patients, plots of land in agriculture experiment.
- Interactions

  A and B interact if A depends on B or vice versa.

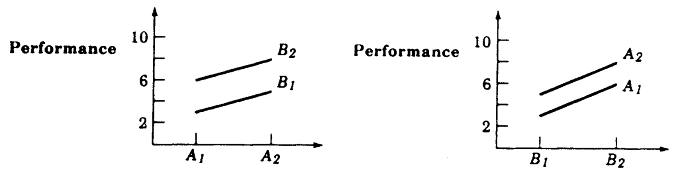
Interactions

TABLE 16.1 Noninteracting Factors

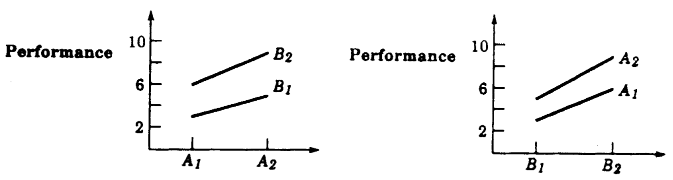
	A_1	A_2
B_1	3	5
B_2	6	8

TABLE 16.2 Interacting Factors

	A_1	A_2
B_1	3	5
B_2	6	9



(a) No Interaction



(b) Interaction

Common Mistakes

- Variation due to errors is ignored
- Important factors are not controlled
- Effects of different factors not isolated
- Simple on-factor-at-a-time design used
- Interactions are ignored
- Too many experiments conducted

Types of Experimental Design

Simple Designs

- Pick a typical configuration
- Vary one factor at a time
- Fix a factor one it is determined best
- Complexity:

$$n = 1 + \sum_{i=1}^{k} (n_i - 1)$$

- Problem: wrong conclusion if there is interaction (suddenly depends on order).
- Not recommended

Full Factorial Design

- Every combination is tried
- Complexity:

$$n = \prod_{i=1}^{k} n_i$$

- Example:
  (3 CPUs)(3 memory levels) x
  (4 disk configs)(3 workloads) x
  (3 edu levels of user) = 324 experim.
- Problem:
  too many experiments
- Solutions:
  reduce number of levels
  reduce number of factors

2^k factorial designs

- restriction to two levels

Fractional Factorial Designs

- Example with 4 factors at 3 level

TABLE 16.3 A Sample Fractional Factorial Design

Experiment Number	CPU	Memory Level	Workload Type	Educational Level
1	68000	512K	Managerial	High school
2	68000	2M	Scientific	Postgraduate
3	68000	8M	Secretarial	College
4	Z80	512K	Scientific	College
5	Z80	2M	Secretarial	High school
6	Z80	8M	Managerial	Postgraduate
7	8086	512K	Secretarial	Postgraduate
8	8086	2M	Managerial	College
9	8086	8M	Scientific	High school

2^k Factorial Designs

Goal:

- Determine the effect of k factors with two levels each

Requirement for reduction to 2 levels

- Factor must be unidirectional

Example 2^2 Factors

Cache Size (kbytes)	Memory Size 4 Mbytes	Memory Size 16 Mbytes
1	15	45
2	25	75

Let us define two variables x_A and x_B as follows:

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

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

The performance y in MIPS can now be regressed on x_A and x_B using a nonlinear regression model of the form

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

Substituting the four observations in the model, we get the following four equations:

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

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

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

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

These four equations can be solved uniquely for the four unknowns. The regression equation is

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

The result is interpreted as follows. The mean performance is 40 MIPS; the effect of memory is 20 MIPS; the effect of cache is 10 MIPS; and the interaction between memory and cache accounts for 5 MIPS. □

Definition:

The q factors are called contrasts

General Approach:

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

$\bar{y}$ mean of responses from all four experiments.

Sum of Squares Total (SST):

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

For a 2^2 design, the variation can be divided into three parts:

$$\text{SST} = 2^2 q_A^2 + 2^2 q_B^2 + 2^2 q_{AB}^2 \qquad (17.1)$$

$$\text{SST} = \text{SSA} + \text{SSB} + \text{SSAB}$$

These parts can be expressed as a fraction; for example,

$$\text{Fraction of variation explained by } A = \frac{\text{SSA}}{\text{SST}}$$

When expressed as a percentage, this fraction provides an easy way to gauge the importance of the factor A.

Example:

TABLE 17.3 Sign Table Method of Calculating Effects in a 2^2 Design

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

Example 17.2 In the case of the memory-cache study

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

$$\text{Total variation} = \sum_{i=1}^{4}(y_i - \bar{y})^2 = (25^2 + 15^2 + 15^2 + 35^2)$$

$$= 2100 = 4 \times 20^2 + 4 \times 10^2 + 4 \times 5^2$$

Thus the total variation is 2100, of which 1600 (76%) can be attributed to memory, 400 (19%) can be attributed to cache, and only 100 (5%) can be attributed to interaction. □

Case Study

Multiprocessor with Network between Memory and Processors:

- Crossbar
- Omega Switching Network

Pattern of Evaluation

- Random Accesses
- Accesses of Matrix Multiply

Factors used/Factors left Constant

TABLE 17.4 Factors Used in the Interconnection Network Study

Symbol	Factor	Level −1	Level 1
A	Type of network	Crossbar	Omega
B	Address pattern used	Random	Matrix

1. Number of processors was fixed at 16.
2. Queued requests were not buffered but blocked.
3. Circuit switching was used instead of packet switching.
4. Random arbitration was used instead of round robin.
5. Infinite interleaving of memory was used so that there was no memory bank contention.

A	B	Response T	Response N	Response R
−1	−1	0.6041	3	1.655
1	−1	0.4220	5	2.378
−1	1	0.7922	2	1.262
1	1	0.4717	4	2.190

Mean Effects for the Interconnection Network Study

Parameter	Mean Estimate T	Mean Estimate N	Mean Estimate R
q_0	0.5725	3.5	1.871
q_A	0.0595	−0.5	−0.145
q_B	−0.1257	1.0	0.413
q_{AB}	−0.0346	0.0	0.051

Parameter	Variation Explained (%) T	Variation Explained (%) N	Variation Explained (%) R
q_0			
q_A	17.2	20	10.9
q_B	77.0	80	87.8
q_{AB}	5.8	0	1.3

Factorial Design with Replication

Example 18.1 The memory-cache experiments were repeated three times each. This resulted in the 12 observations shown in column y in Table 18.1. The analysis is also shown in the table. We sum the individual observations and divide by 3 (the number of replications) to get the sample means $\bar{y}$. The first four columns in the table are sign columns as before. The entries in each of the four columns are multiplied by those in column $\bar{y}$ and the sum is entered under the column. The sums under each column are divided by 4 to give the following effects:

$$q_0 = 41, \quad q_A = 21.5, \quad q_B = 9.5, \quad q_{AB} = 5$$ □

TABLE 18.1 Analysis of a 2^23 Design

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

TABLE 18.2 Computation of Errors in Example 18.2

	Effect				Estimated Response,	Measured Responses			Errors		
	I	A	B	AB							
i	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

Example 18.3 For memory-cache study of Example 18.1:

$$SSY = 15^2 + 18^2 + 12^2 + 45^2 + \cdots + 75^2 + 75^2 + 81^2 = 27{,}204$$
$$SS0 = 2^2 r q_0^2 = 12 \times 41^2 = 20{,}172$$
$$SSA = 2^2 r q_A^2 = 12 \times (21.5)^2 = 5547$$
$$SSB = 2^2 r q_B^2 = 12 \times (9.5)^2 = 1083$$
$$SSAB = 2^2 r q_{AB}^2 = 12 \times 5^2 = 300$$
$$SSE = 27{,}204 - 2^2 \times 3(41^2 + 21.5^2 + 9.5^2 + 5^2) = 102$$
$$SST = SSY - SS0 = 27{,}204 - 20{,}172 = 7032$$

Notice that the SSE computed here is the same as that obtained earlier in Example 18.2. Also, note that the sum of squares add up as follows:

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

Visual Test for Assumptions

- Model errors statistically independent
- Model errors additive
- Errors normally distributed
- Errors have const. Std.Dev.
- Effects of Factors are additive

Independence of Errors (IID)

- compute residuals
- scatter plot

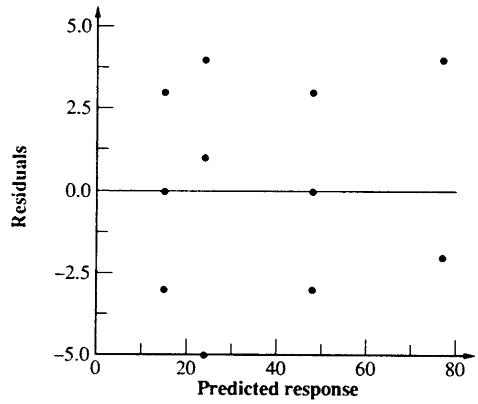
Normal Distribution of Errors

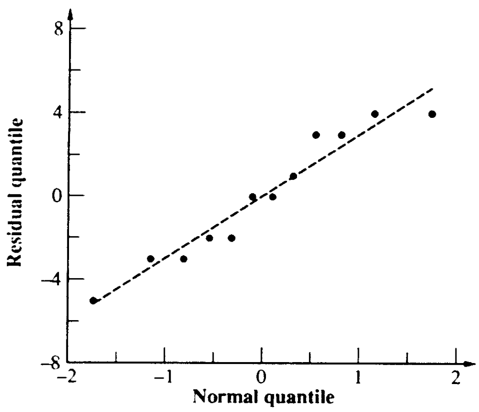
- quantile/quantile plot

Constant Std.Dev. of Errors

- y vs. various levels of factor

Example Plots





Summary of $2^k r$ designs

1. Model: $y_{ij} = q_0 + q_A x_{Ai} + q_B x_{Bi} + q_{AB} x_{Ai} x_{Bi} + \cdots + e_{ij}$
2. 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.
3. Sum of squares: $SSY = \sum_{i=1}^{2^k} \sum_{j=1}^{r} y_{ij}^2$
   $SS0 = 2^k r q_0^2$
   $SST = SSY - SS0$
   $SSj = 2^k r q_j^2,\ j = 1, 2, \ldots, 2^k - 1$
   $SSE = SST - \sum_{j=1}^{2^k-1} SSj$
4. Percentage of y's variation explained by the jth effect: $(SSj/SST) \times 100\%$
5. Standard deviation of errors: $s_e = \sqrt{\frac{SSE}{2^k(r-1)}}$
6. Standard deviation of effects: $s_{q_0} = s_{q_A} = s_{q_B} = s_{q_{AB}} = s_e/\sqrt{2^k r}$
7. Variance of contrast $\sum h_i q_i$, where $\sum h_i = 0$ is $s^2_{\sum h_i q_i} = (s_e^2 \sum h_i^2)/2^k r$
8. Standard deviation of the mean of m future responses:
   $$s_{\hat{y}_p} = s_e \left(\frac{1 + 2^k}{2^k r} + \frac{1}{m} \right)^{1/2}$$
9. Confidence intervals are calculated using $t_{[1-\alpha/2; 2^k(r-1)]}$.
0. Modeling assumptions:
   (a) Errors are IID normal variates with zero mean.
   (b) Errors have the same variance for all values of the predictors.
   (c) Effects of predictors and errors are additive.
1. Visual tests:
   (a) The scatter plot of errors versus predicted responses should not have any trend.
   (b) The normal quantile-quantile plot of errors should be linear.
   (c) Spread of y values in all experiments should be comparable.

   If any test fails or if the ratio y_{max}/y_{min} is large, a multiplicative model should be investigated.