General Full Factorial Design with *k* Factors

Dr. John Mellor-Crummey

Department of Computer Science
Rice University

johnmc@cs.rice.edu

Goals for Today

Understand

- **General Full Factorial Design with *k* Factors**
  - —motivation & model
  - —model properties
  - —estimating model parameters
  - —estimating experimental errors
  - —allocating variation to factors and interactions
  - —analyzing significance of factors and interactions
- **Informal methods**
  - —observation method
  - —ranking method

Full Factorial Design with *k* Factors

- **Motivation**
  - —previous designs we considered had either
    - – only two levels per factor
    - – only two factors
  - —want designs for more than 2 factors, some with more than 2 levels with and without replications
- **General model**
  - —k factors yields 2^k-1 effects
  - —k main effects
  - — $\binom{k}{j}$ j factor interactions, $2 \le j \le k$

Example: Four Factor Paging Study

Symbol	Factor	Level 1	Level 2	Level 3
A	Page Replacement Algorithm	LRUV	FIFO	RAND
D	Deck Arrangement	GROUP	FREQY	ALPHA
P	Problem Program	Small	Medium	Large
M	Memory Pages	24P	20P	16P

Jain reports study from:

Tsao, R.F. and Margolin, B.H. (1971). Multifactor Paging Experiment: II Statistical Methodology. in Freiberger, W., Ed. *Statistical Computer Performance Evaluation*, Academic Press, NY, 135-162.

Example: Model for 4 Factor Design

$$
\begin{aligned}
y_{ijkln} = \mu + \quad & \text{(mean response)} \\
A_i + D_j + P_k + M_l + \quad & \text{(main effects)} \\
\gamma_{ADij} + \gamma_{APik} + \gamma_{AMil} + \gamma_{DPjk} + \gamma_{DMjl} + \gamma_{PMkl} + \quad & \text{(1}^{\text{st}}\text{ order interactions)} \\
\gamma_{ADPijk} + \gamma_{ADMijl} + \gamma_{APMikl} + \gamma_{DPMjkl} + \quad & \text{(2}^{\text{nd}}\text{ order interactions)} \\
\gamma_{ADPMijkl} + \quad & \text{(3}^{\text{rd}}\text{ order interactions)} \\
e_{ijkln} \quad & \text{(error)}
\end{aligned}
$$

$$i = 1,\ldots,a; \quad j = 1,\ldots,d; \quad k = 1,\ldots,p; \quad l = 1,\ldots,m; \quad n = 1,\ldots r$$

n = replications;

(no replications actually considered in following example)

Model Properties for a 4 Factor Design

- **Effects: sums are 0**

$$\sum_i A_i = \sum_j D_j = \sum_k P_k = \sum_l M_l = 0$$

- **Interactions: sum to 0 along each dimension represented**

$$\forall i \quad \sum_{j=1}^{d} \gamma_{ADij} = \sum_{k=1}^{p} \gamma_{APik} = \sum_{l=1}^{m} \gamma_{AMil} = \sum_{j=1}^{d} \gamma_{ADPijk} = \sum_{k=1}^{p} \gamma_{ADPijk} = 0$$

$$\sum_{j=1}^{d} \gamma_{ADMijl} = \sum_{l=1}^{m} \gamma_{ADMijl} = \sum_{k=1}^{p} \gamma_{APMikl} = \sum_{l=1}^{m} \gamma_{APMikl} = 0$$

$$\sum_{j=1}^{d} \gamma_{ADPMijkl} = \sum_{k=1}^{p} \gamma_{ADPMijkl} = \sum_{l=1}^{m} \gamma_{ADPMijkl} = 0$$

similarly for j, k, l dimensions

- **Errors: sum to 0 among all replications of each experiment**

$$\sum_{n=1}^{r} e_{ijkln} = 0, \quad \forall i, j, k, l$$

Example: Four Factor Paging Study

Symbol	Factor	Level 1	Level 2	Level 3
A	Page Replacement Algorithm	LRUV	FIFO	RAND
D	Deck Arrangement	GROUP	FREQY	ALPHA
P	Problem Program	Small	Medium	Large
M	Memory Pages	24P	20P	16P

Data for Paging Study

		GROUP			FREQY			ALPHA		
Algorithm	Program	24P	20P	16P	24P	20P	16P	24P	20P	16P
LRUV	Small	32	48	538	52	244	998	59	536	1348
	Medium	53	81	1901	112	776	3621	121	1879	4639
	Large	142	197	5689	262	2625	10012	980	5698	12880
FIFO	Small	49	67	789	79	390	1373	85	814	1693
	Medium	100	134	3152	164	1255	4912	206	3394	5838
	Large	233	350	9100	458	3688	13531	1633	10022	17117
RAND	Small	62	100	1103	111	480	1782	111	839	2190
	Medium	96	245	2807	237	1502	6007	286	3092	7654
	Large	265	2012	12429	517	4870	18602	1728	8834	23134

Should we use a multiplicative or additive model?

Example: Four Factor Paging Study

Symbol	Factor	Level 1	Level 2	Level 3
A	Page Replacement Algorithm	LRUV	FIFO	RAND
D	Deck Arrangement	GROUP	FREQY	ALPHA
P	Problem Program	Small	Medium	Large
M	Memory Pages	24P	20P	16P

$\log_{10}$ Transformed Data for Paging Study

Algorithm	Program	24P	20P	16P	24P	20P	16P	24P	20P	16P
LRUV	Small	1.51	1.68	2.73	1.72	2.39	3.00	1.77	2.73	3.13
	Medium	1.72	1.91	3.28	2.05	2.89	3.56	2.08	3.27	3.67
	Large	2.15	2.29	3.76	2.42	3.42	4.00	2.99	3.76	4.11
FIFO	Small	1.69	1.83	2.90	1.90	2.59	3.14	1.93	2.91	3.23
	Medium	2.00	2.13	3.50	2.21	3.10	3.69	2.31	3.53	3.77
	Large	2.37	2.54	3.96	2.66	3.57	4.13	3.21	4.00	4.23
RAND	Small	1.79	2.00	3.04	2.05	2.68	3.25	2.05	2.92	3.34
	Medium	1.98	2.39	3.45	2.37	3.18	3.78	2.46	3.49	3.88
	Large	2.42	3.30	4.09	2.71	3.69	4.27	3.24	3.95	4.36

Estimating Model Parameters: 4 Factor Design

- **Organize measured data for two-factor full factorial design as**
  - **a x d x p x m matrix of cells: factor (A,D,P,M) at level (i,j,k,l)**
- **Estimate model parameters**
  - **grand mean**

$$\mu = \bar{y}_{....}$$

  - **main effects**

$$A_i = \bar{y}_{i...} - \bar{y}_{....}$$
$$D_j = \bar{y}_{.j..} - \bar{y}_{....}$$
$$P_k = \bar{y}_{..k.} - \bar{y}_{....}$$
$$M_l = \bar{y}_{...l} - \bar{y}_{....}$$

  - **first order interactions, e.g.**

$$\gamma_{ADij} = \bar{y}_{ij..} - (\bar{y}_{....} + A_i + D_j)$$

$$\gamma_{APik} = \bar{y}_{i.k.} - (\bar{y}_{....} + A_i + P_k)$$

$$\gamma_{DMjl} = \bar{y}_{.j.l} - (\bar{y}_{....} + D_j + M_l)$$

$$\ldots$$

Estimating Model Parameters: 4 Factor Design

- **More model parameter estimates**
  - **second order interactions, e.g.**

$$\gamma_{ADPijk} = \bar{y}_{ijk.} - (\bar{y}_{....} + A_i + D_j + P_k + \gamma_{ADij} + \gamma_{APik} + \gamma_{DPjk})$$

$$\gamma_{ADMijl} = \bar{y}_{ij.l} - (\bar{y}_{....} + A_i + D_j + M_l + \gamma_{ADij} + \gamma_{AMil} + \gamma_{DMjl})$$

$$\ldots$$

  - **third order interaction**

$$\begin{aligned}\gamma_{ADPMijkl} = y_{ijkl} - (\bar{y}_{....} + A_i + D_j + P_k + M_l +{} \\ \gamma_{ADij} + \gamma_{APik} + \gamma_{AMil} + \gamma_{DPjk} + \gamma_{DMjl} + \gamma_{PMkl} +{} \\ \gamma_{ADPijk} + \gamma_{ADMijl} + \gamma_{APMikl} + \gamma_{DPMjkl})\end{aligned}$$

Example: Four Factor Paging Study

Main Effects

	Level		
Factor	1	2	3
A	-0.16	0.02	0.14
D	-0.37	0.07	0.29
P	-0.46	-0.03	0.49
M	-0.69	-0.01	0.70

Degrees of Freedom

Component	DOF
y	81
ybar...	1
y-ybar...	80
Main effects	8
A	2
D	2
P	2
M	2
First order interactions	24
AD	4
AP	4
AM	4
DP	4
DM	4
PM	4
Second-order interactions	32
ADP	8
ADM	8
APM	8
DPM	8
Third-order interactions	16
ADPM	16

Informal Methods I

- **Observation method**
  - —goal: find combination of factor levels yielding best response
  - —how
    - – inspect mean response column
    - – identify
      - high values for HB metric
      - low values for LB metric
    - – unique extreme: associated factor levels give desired combination
    - – multiple extremes: common factor levels provide desired answer

Scheduler Throughput:Observation Method

A	B	C	D	E	Tw	Ti	Tb
-1	-1	-1	-1	1	15.0	25.0	15.2
1	-1	-1	-1	-1	11.0	41.0	3.0
-1	1	-1	-1	-1	25.0	36.0	21.0
1	1	-1	-1	1	10.0	15.7	8.6
-1	-1	1	-1	-1	14.0	63.9	7.5
1	-1	1	-1	1	10.0	13.2	7.5
-1	1	1	-1	1	28.0	36.3	20.2
1	1	1	-1	-1	11.0	23.0	3.0
-1	-1	-1	1	-1	14.0	66.1	6.4
1	-1	-1	1	1	10.0	9.1	8.4
-1	1	-1	1	1	27.0	34.6	15.7
1	1	-1	1	-1	11.0	23.0	3.0
-1	-1	1	1	1	14.0	26.0	12.0
1	-1	1	1	-1	11.0	38.0	2.0
-1	1	1	1	-1	25.0	35.0	17.2
1	1	1	1	1	11.0	22.0	2.0

A=-1 no preemption

B=1 large timeslice

E=1 fairness enabled

Informal Methods II

- **Ranking method: similar to observation method**
  - —rank experiments in decreasing order of response value
  - —observe factor levels that produce consistently good or bad responses

Scheduler Throughput: Ranking Method

A	B	C	D	E	Tw	Ti	Tb
-1	1	1	-1	1	28.0	36.3	20.2
-1	1	-1	1	1	27.0	34.6	15.7
-1	1	-1	-1	-1	25.0	36.0	21.0
-1	1	1	1	-1	25.0	35.0	17.2
-1	-1	-1	-1	1	15.0	25.0	15.2
-1	-1	1	-1	-1	14.0	63.9	7.5
-1	-1	-1	1	-1	14.0	66.1	6.4
-1	-1	1	1	1	14.0	26.0	12.0
1	-1	-1	-1	-1	11.0	41.0	3.0
1	1	1	-1	-1	11.0	23.0	3.0
1	1	-1	1	-1	11.0	23.0	3.0
1	-1	1	1	-1	11.0	38.0	2.0
1	1	1	1	1	11.0	22.0	2.0
1	1	-1	-1	1	10.0	15.7	8.6
1	-1	1	-1	1	10.0	13.2	7.5
1	-1	-1	1	1	10.0	9.1	8.4

A=-1 no preemption good

A=1 preemption bad

B=1 large timeslice good

E=1 fairness enabled

(good if A=-1,B=1)