Recent Advances in Design of Experiment

Dennis Lin
The Pennsylvania State University
DKL5@psu.edu

Design of Experiment

How to collect useful information?

Design Objectives

- Treatment Comparison
- Screening
- Model Building
- Parameter Estimation
- Optimization
- Prediction
- Confirmation
- Discovery (Random Shot)
- etc.

Design Methodology

- Treatment Comparison
- Fractional & Full Factorial Design
- Orthogonal Arrays
- Combinatorics Design
- Coding Theory
- Response Surface Methodology
- ANOVA type Design
- Optimal Design
- Bayesian (Optimal) Design

Design Methodology (Continued)

- Saturated (Minimal Point) Design
- Taguchi Product (Robust) Design
- Mixture Experiment
- Computer Experiment
- Supersaturated Design
- Uniform Design
- MicroArray Design

Design of Experiment (Lin)

- Multiple Response Problems
  - Optimization: Kim and Lin (*JRSS-C*, 2000)
  - Design: Chang, Lo, Lin & Young (*JSPI*, 2001)
- Computer Experiment
  - Beattie and Lin (1999, 2005)
- Dispersion Effect
  - McGrath and Lin (*Technometrics*, 2002)
- Foldover Plan
  - Li and Lin (*Technometrics*, 2003)
- Supersaturated Designs
  - Lin (*Technometrics*, 1993, 1995, 2001) and others
- Uniform Designs
  - Fang, Lin, Winker & Yang (*Technometrics*, 1999)

Dispersion Effect

McGrath and Lin (*JQT*, 2001)
McGrath and Lin (*Technometrics*, 2002)

Injection Molding Process

Y = % shrinkage

X_1 (A) = mold temperature
X_2 (B) = screw speed
X_3 (C) = holding time
X_4 (D) = gate size

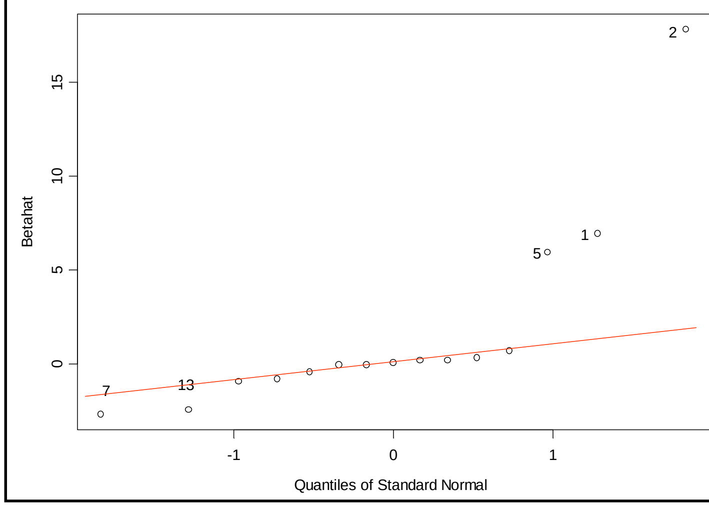
Example: Injection Molding, Montgomery (1990)

	A	B	C	D							E		G	F		
0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	y
1	-1	-1	-1	-1	1	1	1	1	1	1	-1	-1	-1	-1	1	6
1	1	-1	-1	-1	-1	-1	-1	1	1	1	1	1	1	-1	-1	10
1	-1	1	-1	-1	-1	1	1	-1	-1	1	1	1	-1	1	-1	32
1	1	1	-1	-1	1	-1	-1	-1	-1	1	-1	-1	1	1	1	60
1	-1	-1	1	-1	1	-1	1	-1	1	-1	1	-1	1	1	-1	4
1	1	-1	1	-1	-1	1	-1	-1	1	-1	-1	1	-1	1	1	15
1	-1	1	1	-1	-1	-1	1	1	-1	-1	-1	1	1	-1	1	26
1	1	1	1	-1	1	1	-1	1	-1	-1	1	-1	-1	-1	-1	60
1	-1	-1	-1	1	1	1	-1	1	-1	-1	-1	1	1	1	-1	8
1	1	-1	-1	1	-1	-1	1	1	-1	-1	1	-1	-1	1	1	12
1	-1	1	-1	1	-1	1	-1	-1	1	-1	1	-1	1	-1	1	34
1	1	1	-1	1	1	-1	1	-1	1	-1	-1	1	-1	-1	-1	60
1	-1	-1	1	1	1	-1	-1	-1	-1	1	1	1	-1	-1	1	16
1	1	-1	1	1	-1	1	1	-1	-1	1	-1	-1	1	-1	-1	5
1	-1	1	1	1	-1	-1	-1	1	1	1	-1	-1	-1	1	-1	37
1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	52

$\hat{\underset{\sim}{\beta}} \rightarrow$

```
      0       1       2       3       4       5       6       7       8
27.3125 6.9375 17.8125 -0.4375 0.6875 5.9375 -0.8125 -2.6875 -0.9375
      9      10      11      12      13      14      15
-0.0625 -0.0625 0.1875 0.0625 -2.4375 0.1875 0.3125
```

Normal Probability Plot, Injection Molding Example



Traditional Method of Identifying Dispersion Effects in Unreplicated Fractional Factorials

Box and Meyer (1986b) and Montgomery (1990)

1. Identify location effects
2. Fit reduced model and calculate residuals
3. For column d, calculate $F_d^* = ln\frac{s_{d+}^2}{s_{d-}^2}$ (natural log of ratio of sample variance of residuals at $d = \pm 1$.)
4. Box and Meyer point out this statistic is *approximately* normally distributed with mean 0.
5. Montgomery (1990) use a normal probability plot of the F_d^* to identify dispersion effects.

Two Feasible Models

Model I: $\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2 + \hat{\beta}_5 x_5$

Model II: $\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2 + \hat{\beta}_5 x_5 + \hat{\beta}_7 x_7 + \hat{\beta}_{13} x_{13}$

	Model I			Model II		
Column	s_{j+}^2	s_{j-}^2	F_j^*	s_{j+}^2	s_{j-}^2	F_j^*
1	14.43	21.11	-0.38	3.26	2.19	0.40
2	16.11	19.43	-0.19	3.19	2.26	0.34
3	32.44	2.66	$\boxed{2.50}$	2.42	2.58	$\boxed{-0.06}$
4	21.55	12.91	0.51	1.98	2.39	-0.19
5	18.71	16.82	0.11	4.33	1.12	1.36
6	13.55	20.48	-0.41	1.95	1.99	-0.02
7	11.48	7.55	0.42	2.25	3.20	-0.35
8	14.80	18.73	-0.24	2.02	1.42	0.36
9	16.08	19.44	-0.19	2.08	3.35	-0.48
10	22.23	13.30	0.51	2.19	3.25	-0.39
11	17.41	18.05	-0.04	3.17	2.20	0.36
12	22.30	13.23	0.52	2.03	3.41	-0.52
13	12.23	9.73	0.23	1.29	4.16	-1.17
14	15.05	20.41	-0.30	1.16	4.21	-1.29
15	23.76	11.55	0.72	2.96	2.26	0.27

Column 3 is the interaction column of Columns 7 and 13!

F^{ML}, Testing in Presence of Multiple Dispersion Effects

1. Identify location effects.
2. Adapt the location effect model to include exactly the following a terms:
   - The overall mean (β_0),
   - All active location effects and their interactions,
   - The location effects of the columns to be tested for dispersion and their interactions, and
   - The interactions of all of the above.
3. Identify the C_qs, $q = 1, \cdots, m$, the m sets of residuals from rows that are identical for the above columns.
4. Calculate s_q^2 as $\sum_{e_i \in C_q} e_i^2/(n/m - 1)$, i.e. the sample variance of the residuals for each C_q. Each of these sample variance has $d = (n-a)/m$ degrees of freedom.
5. $$F_j^{ML} = \left(\prod_{q:C_q \subset P_j} s_q^2 \prod_{q:C_q \subset M_j} s_q^{-2} \right)^{2/m}$$

Multiple Dispersion Effects (McGrath and Lin, 2001)

Instead of

$$D_E^{BH} = \frac{\sum_{q:C_q \subset P_E} s_q^2}{\sum_{q:C_q \subset M_E} s_q^2} = \frac{s_1^2 + s_2^2}{s_3^2 + s_4^2}$$

use

$$F_E^{ML} = \frac{(\Pi_{q:C_q \subset P_E} s_q^2)^{1/2}}{(\Pi_{q:C_q \subset M_E} s_q^2)^{1/2}} = \frac{(s_1^2 \times s_2^2)^{1/2}}{(s_3^2 \times s_4^2)^{1/2}}$$

Then,

$$R_E^{ML} = \frac{(\Delta_E/\Delta_D\Delta_{DE})^{1/2} \times (\Delta_D\Delta_E\Delta_{DE})^{1/2}}{(\Delta_{DE}/\Delta_D\Delta_E)^{1/2} \times (\Delta_D/\Delta_E\Delta_{DE})^{1/2}} = \boxed{\Delta_E}$$

Summary: Dispersion Effect

- Some Problems
  (see Pan (1999), McGrath and Lin (2001), Brenneman and Nair (2001))
  - unidentified location effects impact dispersion effect identification
  - dispersion effects impact location effect identification
  - multiple dispersion effects very complicated
- Some Solutions
  (see also Bergman and Hynen (1997))
  - Minimal replication (McGrath (2002))
  - Joint confidence regions (McGrath and Lin (2001c))
  - F^{ML} (McGrath and Lin (2002))
- Generalized linear models
  Nelder and Lee (1991), Engel and Huele (1996), McCullagh and Nelder (1989)

Uniform Designs

Fang, Lin, Winker & Yang
(*Technometrics*, 1999)
Fang and Lin
(Handbook of Statistics, Vol 22, 2003)

Uniform Design

A uniform design provides uniformly scatter design points in the experimental domain.

http://www.math.hkbu.edu.hk/UniformDesign

Uniform Design

$\hat{F}_n(x)$ = **Empirical** Cumulative Distribution Function
$F(x)$ = Uniform Cumulative Distribution Function
Find $x = (x_1, x_2, \dots, x_n)$
such that $\hat{F}_n(x)$ is closest to $F(x)$.
Discrepancy

$$D = \left[\int_{\Omega} \left\| \hat{F}_n(x) - F(x) \right\|^p dx \right]^{1/p}$$

- *Wang & Fang (1980)*

The centered L_p-discrepancy is invariant under exchanging coordinates from x to 1-x. Especially, the centered L_2-discrepancy, denoted by CL_2, has the following computation formula:

$$(CL_2(\mathrm{P}))^2$$

$$= \left(\frac{13}{12}\right)^s - \frac{2}{n} \sum_{k=1}^{n} \prod_{i=1}^{s} \left(1 + \frac{1}{2}|x_{ki} - \frac{1}{2}| - \frac{1}{2}|x_{ki} - \frac{1}{2}|^2\right)$$
$$+ \frac{1}{n^2} \sum_{k=1}^{n} \sum_{j=1}^{n} \prod_{i=1}^{s} \left[1 + \frac{1}{2}|x_{ki} - \frac{1}{2}| + \frac{1}{2}|x_{ji} - \frac{1}{2}| - \frac{1}{2}\left|x_{ki} - x_{ji}\right|\right].$$

Uniform Design Example

No.	1	2	3	4	x_1	x_2	x_3	x_4	y
1	11	8	2	10	5.0	40	1.5	60	0.1836
2	9	7	12	8	4.2	35	6.5	50	0.1739
3	8	2	3	2	3.8	10	2.0	20	0.0900
4	10	12	6	4	4.6	60	3.5	30	0.1176
5	1	10	4	7	1.0	50	2.5	45	0.0795
6	2	5	11	3	1.4	25	6.0	25	0.0118
7	4	6	1	5	2.2	30	1.0	35	0.0991
8	7	4	3	12	3.4	20	3.0	70	0.1319
9	6	9	8	1	3.0	45	4.5	15	0.0717
10	3	1	7	9	1.8	5	4.0	55	0.0109
11	5	11	10	11	2.6	55	5.5	65	0.1266
12	12	3	9	6	5.4	15	5.0	40	0.1424

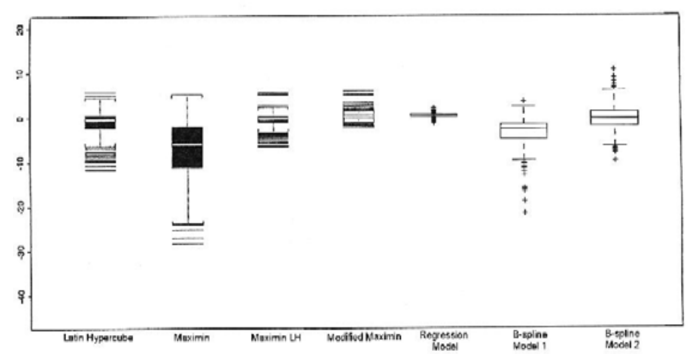
Uniform Design

- *Uniformity*
- *Model Robustness*
- *Flexibility in experimental runs*
- *Flexibility in the number of levels*

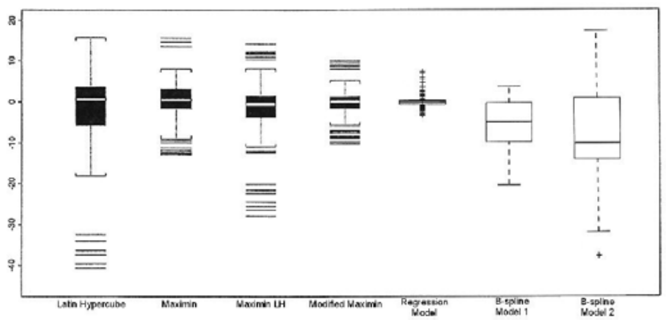
Comparisons
among different designs and models

- *Latin hypercube design*
- *maximin design*
- *maximin Latin hypercube design*
- *modified maximin design*
- *uniform design*

Prediction Errors at 400 Random Samples for Seven Design/Models



Prediction Errors at 256 Corner Points for Seven Design/Models



More on Uniformity Criterion

- **Designs on computer experiments**
- **Factorial designs**
  - design isomorphism
    Ma, Fang and Lin (2001, J. Complexity)
  - consistent with minimum aberration
    Fang and Mukerjee (2000, Biometrika)
  - confounding
    Hikernell and Liu (2000)
  - orthogonality
    Ma, Fang and Lin (2003, JPSI)
  - Hadamard matrix equivalence
    Fang and Ge (2001, accepted by MATH Computation)
- **Supersaturated designs** *(Ma, Fang and Lin, 2000, JPSI)*

Uniform Design: Summary

- Uniformity
- Model Robustness
- Flexibility in experimental runs
- Flexibility in the number of levels

References
- Fang and Lin (2003)
  Handbook of Statistics, Statistics in Industry (Vol.22).
- Fang, Lin, Winker and Zhang
  (Technometrics, 2000)
- Website
  www.math.hkbu.edu.hk/UniformDesign

Optimal Foldover Plan

Li and Lin (Technometrics, 2003)
Li, Ye and Lin (Technometrics, 2003)

Conclusion

The classical wisdom on full foldover is unwise!

A crimp example

- Goal: determine the effect of post crimp stresses on the crimp resistance
- Design: 5=1234, 6=124

  1:crimp height, 2: pro-conditioning thermal shock, 3: dry heat soak, 4: fixture material, 5: thermal shcok life test, 6: discoloration
- Original design: 16-run design
- How do we conduct the next 16-run design?

Notations

- Factors: 1, 2, 3, 4
- Generators: 5=12, 6=134
- Defining relation: I=125=1346=23456
- Word length pattern: W=(0,0,1,1,1)
- Resolution: III
- Foldover plan: γ^f=123456
- WLP of the combined design: W=(0,0,0,1,0)

Resolution, Aberration and *WLP*

- Higher resolution implies less confounding
  - Resolution III designs confound main effects and two-factor interactions
  - Resolution IV designs confound two-factor interactions with some two-factor interactions
- *WLP* (Word Length Pattern) is used to further distinguish designs with same resolution--aberration criterion.

Properties of foldover designs

- Pros
  - Easy to construct
  - Can "separate" effects
  - Combined design still orthogonal
- Cons
  - Large run size

- Our objective:
  Not to compare foldover with other follow-up strategies, but to give the "optimal" foldover among all possible foldover plans

---

- **Question**: In the previous example, in which 5=12, 6=134, is there a "better" foldover plan in terms of *WLP* of combined design?
- **Answer**: γ^*=56 (why???)
  (Trust Dennis!)
- **Objective**: Given a fractional factorial design with economic run size (16 & 32), find its "optimal" foldover plan in terms of the aberration of the combined design.

Core Plan

For any 2^{k-p} design with p generators (G_1,…,G_p), any foldover plan is equivalent to a core foldover plan.

Moreover, for every core foldover plan, there are 2^{k-p} foldover plans that are equivalent to it.

Equivalence of foldover designs

D=

-1	-1	-1	-1
1	-1	-1	1
-1	1	-1	1
1	1	-1	-1
-1	-1	1	1
1	-1	1	-1
-1	1	1	-1
1	1	1	1

γ=123

D	-1	-1	-1	-1
	1	-1	-1	1
	-1	1	-1	1
	1	1	-1	-1
	-1	-1	1	1
	1	-1	1	-1
	-1	1	1	-1
	1	1	1	1
−D	**1**	**1**	**1**	-1
	-1	**1**	**1**	1
	1	**-1**	**1**	1
	-1	**-1**	**1**	-1
	1	**1**	**-1**	1
	-1	**1**	**-1**	-1
	1	**-1**	**-1**	-1
	-1	**-1**	**-1**	1

γ=4

D	-1	-1	-1	-1
	1	-1	-1	1
	-1	1	-1	1
	1	1	-1	-1
	-1	-1	1	1
	1	-1	1	-1
	-1	1	1	-1
	1	1	1	1
D_4	-1	-1	-1	**1**
	1	-1	-1	**-1**
	-1	1	-1	**-1**
	1	1	-1	**1**
	-1	-1	1	**-1**
	1	-1	1	**1**
	-1	1	1	**1**
	1	1	1	**-1**

Core foldover plans

- Core foldover plan: consists only of the generated factors
- Example: $\gamma = 5$
- ***Theorem***: Any foldover plan is equivalent to a core foldover plan.
- Examples
  - $\gamma^f = 123456 \Leftrightarrow \gamma = 5$
  - $\gamma^* = 1 \Leftrightarrow \gamma^* = 56$
- There are essentially 2^p choices.

Foldover of resolution III designs

Table 3. Word Length Pattern Comparisons on Selected Foldover Resolution III Designs

Design k-p	Initial $W(\boldsymbol{d})$	$W(\boldsymbol{D}(\gamma^f))$ of full-foldover	$W(\boldsymbol{D}(\gamma^*))$ of optimal foldover	Optimal foldover plan (γ*)
6-2.2	(1 1 1 0)	(0 1 0 0 0)	(0 0 1 0 0)	56
7-3.2	(2 3 2 0 0)	(0 3 0 0 0)	(0 1 2 0 0)	567
8-4.2	(3 7 4 0 1)	(0 7 0 0 0)	(0 3 4 0 0)	5678
8-4.4	(4 6 4 0 0)	(0 6 0 0 0)	(0 3 4 0 0)	567
9-5.1	(4 14 8 0 4)	(0 14 0 0 0)	(0 6 8 0 0)	5678
7-2.5	(1 1 0 0 1)	(0 1 0 0 0)	(0 0 0 0 1)	67
8-3.5	(1 2 3 1 0)	(0 2 0 1 0)	(0 0 2 1 0)	678
9-4.6	(1 5 6 2 1)	(0 5 0 2 0)	(0 1 4 2 0)	6789
9-4.7	(1 7 4 0 3)	(0 7 0 0 0)	(0 3 2 0 2)	67,…,69
9-4.8	(2 3 6 4 0)	(0 3 0 4 0)	(0 1 4 2 0)	6,786,789,679
10-5.5	(1 14 7 0 7)	(0 14 0 0 0)	(0 6 4 0 4)	678, …, 69<u>10</u>
11-6.6	(3 13 19 11 9)	(0 13 0 11 0)	(0 5 12 7 4)	678

A Typical Example

- Factors: 1, 2, 3, 4
- Generators: 5=12, 6=134
- Defining relation: I=125=1346=23456
- Word length pattern: W=$(0,0,1,1,1,0)_{III}$
- Full Foldover plan: γ^f=123456 [γ=5]
- WLP of the combined design: W=$(0,0,0,1,0,0)_{IV}$
- Optimal Foldover plan: γ*=56
- WLP of the combined design: W=$(0,0,0,0,1,0)_{V}$

Foldover of resolution IV designs

Li & Lin (2003)

Table 4. Word Length Pattern Comparisons on Selected Foldover Resolution IV Designs

Design k-p	Initial $W(\boldsymbol{d})$	$W(\boldsymbol{D}(\gamma^f))$ of full-foldover	$W(\boldsymbol{D}(\gamma^*))$ of optimal foldover	Optimal foldover plan (γ*)
5-1.2	(0 1 0)	(0 1 0 0 0)	full factorial	5
6-2.1	(0 3 0 0)	(0 3 0 0 0)	(0 1 0 0 0)	5,56,6
7-3.1	(0 7 0 0 0)	(0 7 0 0 0)	(0 3 0 0 0)	5, …, 7
8-4.1	(0 14 0 0 0)	(0 14 0 0 0)	(0 6 0 0 0)	56, …, 78
7-2.1	(0 1 2 0 0)	(0 1 0 0 0)	(0 0 1 0 0)	6, 7
8-3.1	(0 3 4 0 0)	(0 3 0 0 0)	(0 1 2 0 0)	6, …, 78
9-4.3	(0 9 0 6 0)	(0 9 0 6 0)	(0 3 0 4 0)	678, …, 789
9-4.5	(0 14 0 0 0)	(0 14 0 0 0)	(0 6 0 0 0)	67, …, 89
10-5.1	(0 10 16 0 0)	(0 10 0 0 0)	(0 4 8 0 0)	67, …, 9<u>10</u>
11-6.2	(0 26 0 24 0)	(0 26 0 24 0)	(0 10 0 16 0)	78<u>10</u>,…, 89<u>10</u> <u>11</u>

Major results

- For most designs there exist better foldover plans than the full foldover plan
  - In 52 out of 77 designs have better foldover plans than full-foldover plans
  - Most (42) foldover plans are new
- Almost all minimum aberration designs have better foldover plans

Summary *(Li & Lin, 2003)*

- Proposed a computer search method to construct optimal foldover plans that minimizes the WLP of the combined design.
- Tabulated optimal foldover plans for commonly-used 16-run and 32-run designs.
- Investigated optimal and full foldover plans by focusing on core foldover plans.
- Demonstrated that there exists a unique group of equivalent foldover plans.
- See also some work of Mee.
- Theory on optimal foldover—Fang, Lin and Qin(2003)

Theoretical Support on Optimal Flodover
(via Discrepancy)

Fang, Lin and Qin (2003)

Discrepancy for Design D

$$(\mathrm{CD}(\mathscr{P}))^2 = \left(\frac{13}{12}\right)^s - \frac{1}{n}\sum_{k=1}^{n}\prod_{i=1}^{s}\left(1 - \frac{1}{2}\left|x_{ki} - \frac{1}{2}\right| - \frac{1}{2}\left|x_{ki} - \frac{1}{2}\right|^2\right)$$
$$+\frac{1}{n^2}\sum_{k=1}^{n}\sum_{j=1}^{n}\prod_{i=1}^{s}\left(1 + \frac{1}{2}\left|x_{ki} - \frac{1}{2}\right| + \frac{1}{2}\left|x_{ji} - \frac{1}{2}\right| - \frac{1}{2}|x_{ki} - x_{ji}|\right)$$

Discrepancy for combined design under foldover plan γ

$$(\mathrm{CD}(T(\gamma)))^2 = c + \frac{1}{2n^2}\sum_{k=1}^{n}\sum_{j=1}^{n}\prod_{i=1}^{s}\left(\frac{5}{2} - \frac{1}{2}|x_{ki} - x_{ji}| - \frac{1}{2}|x_{ki} + (-1)^{d_i+1}x_{ji} - \delta_i|\right)$$

This can be shown to be minimum among all designs of same size.

Future work

- Optimal foldover designs of non-regular designs
  - Generalized resolution criterion (Deng and Tang, 1999)
  - Allows "fractional" word length
  - Miller and Sitter (2001)
  - Li, Ye and Lin (2003)
- Optimal semi-foldover designs built on Mee and Peralta (2000) and this work
- Fold over high level designs
- Optimal follow-up experiment, in general.

Supersaturated Designs

Lin (UTK Technical Report, 1991)
Lin (*Technometrics*, 1993, 1995, 2001)
Lin (Handbook of Statistics, Vol 22, 2003)
and others

Supersaturated Design Example

Half Fraction of William's (1968) Data

	Factor																								
Run	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	y
1	+	+	+	-	-	-	+	+	+	+	+	-	+	-	-	+	+	-	-	+	-	-	-	+	133
2	+	-	-	-	-	-	+	+	+	-	-	-	+	+	+	+	-	+	-	-	+	+	-	-	62
3	+	+	-	+	+	-	-	-	-	+	-	+	+	+	+	+	+	-	-	-	-	+	+	-	45
4	+	+	-	+	-	+	-	-	-	+	+	-	+	-	+	+	-	+	+	+	-	-	-	-	52
5	-	-	+	+	+	+	-	+	+	-	-	-	+	-	+	+	+	-	-	+	-	+	+	+	56
6	-	-	+	+	+	+	+	-	+	+	+	-	-	+	+	-	+	+	+	+	+	+	-	-	47
7	-	-	-	-	+	-	-	+	-	+	-	+	+	+	-	+	+	+	+	+	+	-	-	+	88
8	-	+	+	-	-	+	-	+	-	+	-	-	-	-	-	-	-	-	+	-	+	+	+	-	193
9	-	-	-	-	-	+	+	-	-	-	+	+	-	-	+	-	+	+	-	-	-	-	+	+	32
10	+	+	+	+	-	+	+	+	-	-	-	+	-	+	+	-	+	-	+	-	+	-	-	+	53
11	-	+	-	+	+	-	-	+	+	-	+	-	-	+	-	-	-	+	+	-	-	-	+	+	276
12	+	-	-	-	+	+	+	-	+	+	+	+	+	-	-	+	-	-	+	-	+	+	+	+	145
13	+	+	+	+	+	-	+	-	+	-	-	+	-	-	-	-	-	+	-	+	+	-	+	-	130
14	-	-	+	-	-	-	-	-	-	-	+	+	-	+	-	-	-	-	-	+	-	+	-	-	127

Lin (1993, *Technometrics*)

SUPERSATURATED DESIGN

How can we study k parameters
with n(<k) observations (experiments)?

A situation for using supersaturated design:

- *A Small number of run is desired*
- *The number of potential factors is large*
- *Only a few active factors*

Design Criteria

Supersaturated Design

- Booth and Cox (1962): $E(s^2)$
- Wu (1993): Extension of classical optimalities (D_f, A_f etc)
- Deng and Lin (1994): 8 criteria
- Deng, Lin and Wang (1996): B-cirterion
- Deng, Lin and Wang (1994): resolution rank
- Balkin and Lin (1997):
  Graphical Comparison (Harmonic mean of eigens)

New Age SSD Criteria

- $E(f_{NOD})$
- Discrete Discrepancy
- Minimum Generalized Aberration (MGA)
- Generalized Minimum Aberration (GMA)
- Indicator Function
- Relationships/Connections Among criteria
- Probability of Correct Search

Design Construction

Supersaturated Design

- *Half Fraction of Hadamard Matrix*
- *Random Combined Design (Taguchi, 1986)*
- *Algorithmic Approach*
- *Combinatorial Approach*
- *Optimal Supersaturated Design*
- *Others*

Supersaturated Design From Hadamard Matrix of Order 12
(Using 11 as the branching column)

Run	Factors										
No.	I	1	2	3	4	5	6	7	8	9	10	(11)
1	+	+	+	-	+	+	+	-	-	-	+	-
2	+	+	-	+	+	+	-	-	-	+	-	+
3	+	-	+	+	+	-	-	-	+	-	+	+
4	+	+	+	+	-	-	-	+	-	+	+	-
5	+	+	+	-	-	-	+	-	+	+	-	+
6	+	+	-	-	-	+	-	+	+	-	+	+
7	+	-	-	-	+	-	+	+	-	+	+	+
8	+	-	-	+	-	+	+	-	+	+	+	-
9	+	-	+	-	+	+	-	+	+	+	-	-
10	+	+	-	+	+	-	+	+	+	-	-	-
11	+	-	+	+	-	+	+	+	-	-	-	+
12	+	-	-	-	-	-	-	-	-	-	-	-

Supersaturated Design Example

Half Fraction of William's (1968) Data

	Factor																								
Run	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	y
1	+	+	+	-	-	-	+	+	+	+	+	-	+	-	-	+	+	-	-	+	-	-	-	+	133
2	+	-	-	-	-	-	+	+	+	-	-	-	+	+	+	+	-	+	-	-	+	+	-	-	62
3	+	+	-	+	+	-	-	-	-	+	-	+	+	+	+	+	+	-	-	-	-	+	+	-	45
4	+	+	-	+	-	+	-	-	-	+	+	-	+	-	+	+	-	+	+	+	-	-	-	-	52
5	-	-	+	+	+	+	-	+	+	-	-	-	+	-	+	+	+	-	-	+	-	+	+	+	56
6	-	-	+	+	+	+	+	-	+	+	+	-	-	+	+	-	+	+	+	+	+	+	-	-	47
7	-	-	-	-	+	-	-	+	-	+	-	+	+	+	-	+	+	+	+	+	+	-	-	+	88
8	-	+	+	-	-	+	-	+	-	+	-	-	-	-	-	-	-	-	+	-	+	+	+	-	193
9	-	-	-	-	-	+	+	-	-	-	+	+	-	-	+	-	+	+	-	-	-	-	+	+	32
10	+	+	+	+	-	+	+	+	-	-	-	+	-	+	+	-	+	-	+	-	+	-	-	+	53
11	-	+	-	+	+	-	-	+	+	-	+	-	-	+	-	-	-	+	+	-	-	-	+	+	276
12	+	-	-	-	+	+	+	-	+	+	+	+	+	-	-	+	-	-	+	-	+	+	+	+	145
13	+	+	+	+	+	-	+	-	+	-	-	+	-	-	-	-	-	+	-	+	+	-	+	-	130
14	-	-	+	-	-	-	-	-	-	-	+	+	-	+	-	-	-	-	-	+	-	+	-	-	127

Lin (1993, *Technometrics*)

Half Fraction Hadamard Matrix

$(n, k) = (2t, 4t - 2)$

$\updownarrow$

Balanced Incomplete Block Design

$$\upsilon = 2t - 1$$
$$b = 4t - 2$$
$$\gamma = 2t - 2$$
$$k = t - 2$$

Hedayat & Wallis (1978)

- $ave(s^2) = n^2 / (2n - 3)$
  proved to be $E(s^2)$-optimal!
- Non-isormorphic class exist!

Half Fraction Hadamard Matrix

$(n, k) = (2t, 4t - 2)$

$\updownarrow$

Coding

Binary Code

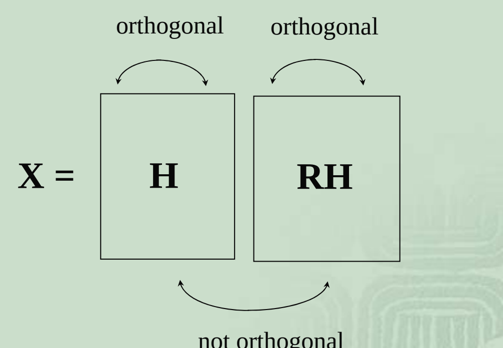
length $n = 2t$

weight $\omega = t$

distance d

$$\frac{n}{3} \le d \le \frac{2n}{3}$$
$$if\ |\gamma| \le 1/3$$

- Find $A[n, d, \omega]$: maximum number of codewords.



Thus Permute rows of RH to minimize $E(s^2)$, say.

Algorithmic

Supersaturated Design

Lin (1991, 1995): Pairwise Optimality

Nguyen (1996): Exchange Algorithm

Li and Wu (1997): Columnwise and Pairwise Algorithm

Church (1993): Projection Properties

$$U \oplus L = \begin{array}{c} \text{UD} \\ \begin{bmatrix} 1 & 1 \\ 2 & 7 \\ 3 & 3 \\ 4 & 9 \\ 5 & 5 \\ 6 & 6 \\ 7 & 2 \\ 8 & 8 \\ 9 & 4 \end{bmatrix} \end{array} \oplus \begin{array}{c} \text{OD} \\ \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 1 \\ 0 & 2 & 2 & 2 \\ 1 & 0 & 1 & 2 \\ 1 & 1 & 2 & 0 \\ 1 & 2 & 0 & 1 \\ 2 & 0 & 2 & 1 \\ 2 & 1 & 0 & 2 \\ 2 & 2 & 1 & 0 \end{bmatrix} \end{array} = X = \begin{array}{c} \text{SSD} \\ \left[\begin{array}{cccc|cccc} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 1 & 2 & 0 & 2 & 1 \\ 0 & 2 & 2 & 2 & 0 & 2 & 2 & 2 \\ 1 & 0 & 1 & 2 & 2 & 2 & 1 & 0 \\ 1 & 1 & 2 & 0 & 1 & 1 & 2 & 0 \\ 1 & 2 & 0 & 1 & 1 & 2 & 0 & 1 \\ 2 & 0 & 2 & 1 & 0 & 1 & 1 & 1 \\ 2 & 1 & 0 & 2 & 2 & 1 & 0 & 2 \\ 2 & 2 & 1 & 0 & 1 & 0 & 1 & 2 \end{array}\right] \end{array}$$

Fang, Lin & Ma (2000)

Data Analysis Methods

Supersaturated Design

- Pick-the-Winner
- Graphical Approach
- "PARC" (Practical Accumulation Record Computation)
- Compact Two-Sample Test
- Forward Selection
- Ridge Regression
- Normal Plot

Design Analysis

Supersaturated Design

- *Classical Approaches*
- *Adjusted p-value*
  (Westfall, Young & Lin, Statistica Sinica, 1998)
- *Bayesian Approach*
  (Beattie, Fong & Lin, Technometrics, 2002)
- *Penalized Least Squares*
  (Li & Lin, 2002)

SSD: Looking Ahead

Supersaturated Design

- SSD is much more mature than ever
- Nano-Manufacturing Applications
- Micro-Array Design and Analysis
- Computer Experiment: Model Building (using SSD)
- Higher (and Mixed) Level SSD
- Spotlight Interaction Effects (Lin, 1998, QE)
- Combination Designs: Rotated FFD & SSD

Computer Experiment

Beattie and Lin (1999 & 2005)

Computer Experiment

- Expensive simulation
- When Monte Carlo study is infeasible, how to run simulation?
- Latin Hypercube

Goals—Computer Experiment

- Confirmation
- Sensitivity Analysis
- Empirical Model Building
- Optimization
- Model Validation
- High Dimension Integration

Statistics vs. Engineering Models

$$y = f(x,\theta) + \varepsilon$$

Statistical Model

$$y=\beta_0+\Sigma\beta_i x_i+\Sigma\beta_{ij} x_i x_j+\varepsilon$$

Irrelevant Issues

- Replicates
- Blocking
- Randomization

Question: How can a computer experiment be run in an efficient manner?

Current Approaches to Experimental Design

- Geometric (Frequentist) Designs
  - Full and Fractional Factorial Designs
  - Other Traditional Designs
  - Latin Hypercube Designs (McKay, Beckman, and Conover (1979))
- Computer-Generated (Bayesian) Designs
  - Maximin Distance Designs (Johnson, Moore, and Ylvisaker (1990))
- Combination Designs (Computer-Generated Geometric)
  - Maximin Latin Hypercube Designs (Morris and Mitchell (1992))
  - Orthogonal Array-based LHs (Tang (1993), Owen (1992))
  - Rotated Factorial Designs (Beattie and Lin (1999))

Rotated Factorial Designs

- Computer experiments are gaining in popularity
  - main research area of the next 10 years
- Rotated factorial designs
  - good factorial design properties (orthogonality and structure)
  - good Latin hypercube properties (unique and equally-spaced projections)
  - easy to construct
  - comparable by Bayesian criteria
  - very suitable for computer experiments

$$D = X \cdot V$$

desirable design — factorial design — rotated matrix

$$\begin{bmatrix} 1 & 1 \\ 2 & 1 \\ \vdots & \vdots \\ p & 1 \\ 1 & 2 \\ 2 & 2 \\ \vdots & \vdots \\ p & 2 \\ \vdots & \vdots \\ 1 & p \\ 2 & p \\ \vdots & \vdots \\ p & p \end{bmatrix}_{p^2 \times d} \qquad \begin{bmatrix} \upsilon_1 & \upsilon_3 \\ \upsilon_2 & \upsilon_4 \end{bmatrix}_{d \times d}$$

Beattie & Lin (1998): Rotating Full Factorials

$$D = X \cdot V$$

desirable design — fractional factorial design (mostly two-level) — rotated matrix

$$\begin{bmatrix} S_1 & & & & & \\ & S_2 & & & & \\ & & \ddots & & & \\ & & & S_i & & \\ & & & & \ddots & \\ & & & & & S_m \end{bmatrix}_{n \times n}$$

Bursztyn & Steinberg (2002): Rotating in Groups

Now,
Put these two ideas together!

- Grouping all design columns into groups,
- each forms a full factorial design,
- then rotate each group (in block).

Steinberg and Lin (*Biometrika*, 2006)

Multiple Response Problems

Optimization: Kim and Lin (*JRSS-C*, 2000)

Design: Chang, Lo, Lin & Young (*JSPI*, 2001)

CGA Study

Data:

Central Composite Design in x_1, x_2 & x_3, with three outputs, y_1, y_2 & y_3 and three (3) replicates.

Fitted "Location" Models

$\hat{y}_{\mu_1}(x) = 4.95 + 0.82x_1 - 0.45x_2 - 0.15x_1^2 + 0.28\,x_2^2 - 0.11x_1x_2 + 0.07x_1x_3$ ($R^2 = 0.91$)

$\hat{y}_{\mu_2}(x) = 0.46 + 0.13x_1 - 0.06x_2 + 0.05\,x_3 - 0.07x_1^2 - 0.04\,x_3^2$ ($R^2 = 0.87$)

$\hat{y}_{\mu_3}(x) = 28.36 - 1.48x_1 + 2.33x_3 - 0.15x_1^2 - 1.42\,x_2^2 - 0.71x_1x_3$ ($R^2 = 0.12$)

***Objective**:*

Find x such that all y_1, y_2 & y_3 are simultaneously "optimized" .*

CGA Study

Fitted "Dispersion" Models

$\hat{y}_{\sigma 1}(\mathbf{x}) = 0.06 + 0.11x_2 + 0.06x_3 + 0.12x_1^2 + 0.11x_3^2 - 0.10x_1x_3 + 0.05x_2x_3$ ($R^2 = 0.84$)

$\hat{y}_{\sigma 2}(\mathbf{x}) = 0.02 - 0.01x_1 + 0.01x_2 - 0.01x_3 + 0.02x_3^2 - 0.01x_1x_3 + 0.02x_2x_3$ ($R^2 = 0.83$)

$\hat{y}_{\sigma 3}(\mathbf{x}) = 6.08 - 1.53x_1 + 0.50x_2 + 4.85x_3 + 2.26\,x_2^2 - 0.65x_1x_3 - 0.67\,x_1x_2x_3$ ($R^2 = 0.95$)

Fitted "Location" Models

$\hat{y}_{\mu_1}(x) = 4.95 + 0.82x_1 - 0.45x_2 - 0.15x_1^2 + 0.28\,x_2^2 - 0.11x_1x_2 + 0.07x_1x_3$ ($R^2 = 0.91$)

$\hat{y}_{\mu_2}(x) = 0.46 + 0.13x_1 - 0.06x_2 + 0.05\,x_3 - 0.07x_1^2 - 0.04\,x_3^2$ ($R^2 = 0.87$)

$\hat{y}_{\mu_3}(x) = 28.36 - 1.48x_1 + 2.33x_3 - 0.15x_1^2 - 1.42\,x_2^2 - 0.71x_1x_3$ ($R^2 = 0.12$)

***Objective**:*

Find x such that all y_1, y_2 & y_3 are simultaneously "optimized" , when both location & dispersion responses are under concerned!.*

MRS Optimization : Approaches

- Priority–based Approach
- Desirability Function Approach*
- Generalized Distance Approach*
- Loss Function Approach*

* **dimensionality reduction strategy**

MRS Optimization : Priority – based Approach

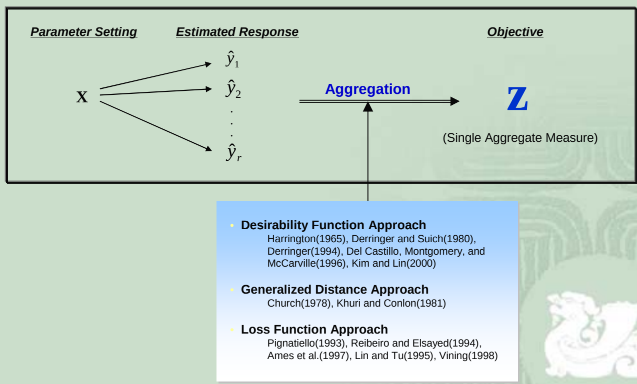
- Primary response vs. Secondary responses
- Framework

$$\begin{aligned} \underset{\mathbf{x}}{\text{Optimize}} \quad & \text{Primary response} \\ \text{s.t} \quad & \text{Requirements for secondary responses} \\ & \mathbf{x} \in \Omega \end{aligned}$$

- Related Work

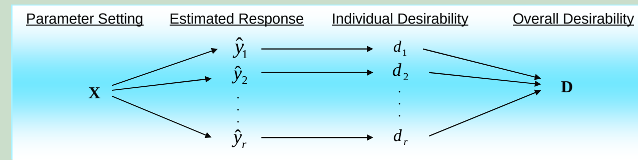
Hoerl (1959)	Del Castillo and Montgomery (1993)
Myers and Carter(1973)	Copeland and Nelson (1996)
Biles (1975)	Semple (1997)
Vining and Myers (1990)	Del Castillo, Fan, and Semple (1999)

Framework of Dimensionality Reduction Strategy



MRS Optimization : Desirability Function Approach

- Framework

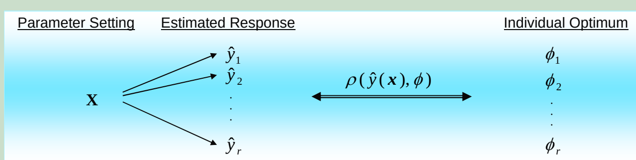


- Find **X*** to Maximize D
- Related work

Harrington (1965)	Kim and Lin (1998)
Derringer and Suich (1980)	Kim and Lin (2000)
Derringer (1994)	Del Castillo, Montgomery, and McCarville (1996)
Goik, Liddy, and Taam (1994)	

MRS Optimization : Generalized Distance Approach

- Framework



- $\rho(\hat{y}(\mathbf{x}), \phi)$= Distance between $\hat{y}(\mathbf{x})$ and ϕ
- Find $\mathbf{x}^*$ to Minimize $\rho(\hat{y}(\mathbf{x}), \phi)$
- Related work
  - Church (1978)
  - Khuri and Conlon (1981)

MRS Optimization : Generalized Distance Approach (cont'd)

- Khuri and Conlon (1981)
  - Distance of Estimated Responses from Estimated "Ideal" Optimum

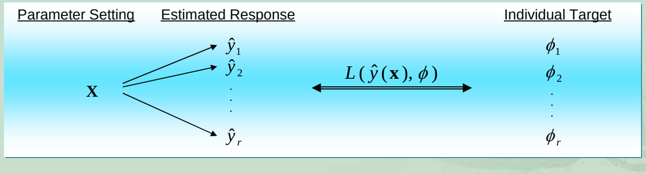
$$\rho[\hat{\mathbf{y}}(\mathbf{x}), \phi] = \left[(\hat{\mathbf{y}}(\mathbf{x}) - \phi)' \hat{\Sigma}^{-1} (\hat{\mathbf{y}}(\mathbf{x}) - \phi) / \mathbf{z}'(\mathbf{x})(\mathbf{X}'\mathbf{X})^{-1} \mathbf{z}(\mathbf{x})\right]^{1/2},$$

where $\phi = [\phi_1, \phi_2, ..., \phi_r]'$ is the ideal optimum,
$\hat{\Sigma}$ is the estimator of the common variance-covariance matrix of the random errors $(\varepsilon_1, \varepsilon_2, ..., \varepsilon_r)$,
X is the design matrix, and
z(x) is a column vector of the input variables of the given model.

  - Assume All Response Functions
    - Depend on the same set of input variables.
    - Are of the same form.

MRS Optimization : Loss Function Approach

- Framework



- $L(\hat{y}(\mathbf{x}),\phi) = [\hat{y}(\mathbf{x})-\phi]' C\, [\hat{y}(\mathbf{x})-\phi]$ (Multivariate squared error loss)
- Find $\mathbf{x}^*$ to minimize $E(L)$
- Related Work

Pignatiello (1993)
Ames et al. (1997)
Vining (1998)
Ribeiro and Elsayed (1994)
Lin and Tu (1995)

Example : Colloidal Gas Aphrons (CGA) Study

- Characterization of CGA Properties (Jauregi et al. 1997)
- Responses : Stability (y_1, LTB), Volumetric Ratio (y_2, STB), Temperature (y_3, NTB)
- Input Variables : Concentration of Surfactant (x_1), Concentration of Salt (x_2), Time of Stirring (x_3)
- Design : CCD with 8 Factorial Points*, 6 Axial Points*, and a Center Point**

(* Replicated twice, ** Replicated 6 times)

Example : CGA Study (continued)

- Fitted "Location" Models

$\hat{y}_{\mu_1}(x)$ = 4.95 + 0.82x_1 – 0.45x_2 – 0.15x_1^2 + 0.28 x_2^2 - 0.11x_1x_2 + 0.07x_1x_3 (R^2 = 0.91)

$\hat{y}_{\mu_2}(x)$ = 0.46 + 0.13x_1 – 0.06x_2 + 0.05 x_3 – 0.07x_1^2 - 0.04 x_3^2 (R^2 = 0.87)

$\hat{y}_{\mu_3}(x)$ = 28.36 – 1.48x_1 + 2.33x_3 – 0.15x_1^2 – 1.42 x_2^2 - 0.71x_1x_3 (R^2 = 0.12)

- Linear Desirability Functions (for simplicity)
- Derringer and Suich (DS) Method :

$$\underset{\mathbf{x}}{\text{Maximize}} \quad \sqrt[3]{d_{\mu 1}(\hat{y}_{\mu 1})\, d_{\mu 2}(\hat{y}_{\mu 2})\, d_{\mu 3}(\hat{y}_{\mu 3})}$$

$$\text{s.t} \quad -1 \le x_i \le 1 \qquad (i = 1,2,3)$$

Example : CGA Study (continued)

	Responses		
	y_1	y_2	y_3
Bounds and Target			
$y_{\mu_j}^{\min}$, $y_{\sigma_j}^{\min}$	3.00, 0.00	0.10, 0.00	15.00, 1.00
$y_{\mu_j}^{\max}$, $y_{\sigma_j}^{\max}$	7.00, 0.10	0.60, 0.10	45.00, 2.00
T_{μ_j} , T_{σ_j}	7.00, 0.00	0.10, 0.00	30.00, 1.00
Optimization Results			
DS Method	$\mathbf{x}_{DS}^*$ = (-1.00, -1.00, -1.00)		
$\hat{y}_{\mu_j}(\mathbf{x}_{DS}^*)$, $\hat{y}_{\sigma_j}(\mathbf{x}_{DS}^*)$	4.66, *0.06*	0.24, *0.08*	25.38, *4.54*
$d_{\mu_j}(\hat{y}_{\mu_j}(\mathbf{x}_{DS}^*))$, $d_{\sigma_j}(\hat{y}_{\sigma_j}(\mathbf{x}_{DS}^*))$	0.41, *0.41*	0.72, *0.23*	0.69, *0.00*

† The $\hat{y}_{\sigma j}$ and $d_{\sigma j}(\hat{y}_{\sigma j})$ values for the standard deviation responses are computed *a posteriori* at the given $\mathbf{x}_{DS}^*$, and are written in italic.

- Fitted "Dispersion" Models

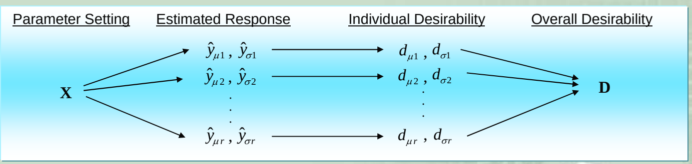
$\hat{y}_{\sigma 1}(\mathbf{x})$ = 0.06 + 0.11x_2 + 0.06x_3 + 0.12x_1^2 + 0.11x_3^2 - 0.10x_1x_3 + 0.05x_2x_3 (R^2 = 0.84)

$\hat{y}_{\sigma 2}(\mathbf{x})$ = 0.02 - 0.01x_1 + 0.01x_2 - 0.01x_3 + 0.02x_3^2 – 0.01x_1x_3 + 0.02x_2x_3 (R^2 = 0.83)

$\hat{y}_{\sigma 3}(\mathbf{x})$ = 6.08 – 1.53x_1 + 0.50x_2 + 4.85x_3 + 2.26 x_2^2 - 0.65x_1x_3 - 0.67 $x_1x_2x_3$ (R^2 = 0.95)

Proposed Approach* : Framework

- Consideration of Both Location and Dispersion Effects
- "Maximining" Desirability Functions
- Framework



$\mathbf{D} = Minimum\ \{d_{\mu 1}, \cdots, d_{\mu r}, d_{\sigma 1}, \cdots, d_{\sigma r}\}$ ➡ Find **X*** to Maximize **D**.

* Co-work with Kwang-Jae Kim

Proposed Approach : Formulation

$$\underset{x}{\text{Maximize}} \quad \lambda$$

$$\text{subject to} \quad d_{\mu_j}(\hat{y}_{\mu_j}(\boldsymbol{x})) \geq \lambda, \quad j = 1, 2, ..., r,$$

$$d_{\sigma_j}(\hat{y}_{\sigma_j}(\boldsymbol{x})) \geq \lambda, \quad j = 1, 2, ..., r,$$

$$\mathbf{x} \in \Omega\ .$$

Example : CGA Study - Revisited

	Responses		
	y_1	y_2	y_3
Bounds and Target			
$y_{\mu_j}^{\min}$, $y_{\sigma_j}^{\min}$	3.00, 0.00	0.10, 0.00	15.00, 1.00
$y_{\mu_j}^{\max}$, $y_{\sigma_j}^{\max}$	7.00, 0.10	0.60, 0.10	45.00, 2.00
T_{μ_j} , T_{σ_j}	7.00, 0.00	0.10, 0.00	30.00, 1.00
Optimization Results			
DS Method	$\boldsymbol{x}_{DS}^* = (-1.00, -1.00, -1.00)$		
$\hat{y}_{\mu_j}(\mathbf{x}_{DS}^*)$, $\hat{y}_{\sigma_j}(\mathbf{x}_{DS}^*)$	4.66, *0.06*	0.24, *0.08*	25.38, *4.54*
$d_{\mu_j}(\hat{y}_{\mu_j}(\mathbf{x}_{DS}^*))$, $d_{\sigma_j}(\hat{y}_{\sigma_j}(\mathbf{x}_{DS}^*))$	0.41, *0.41*	0.72, *0.23*	0.69, *0.00*
Proposed Method	$\boldsymbol{x}_p^* = (-0.21, -0.40, -1.00)$		
$\hat{y}_{\mu_j}(\mathbf{x}_P^*)$, $\hat{y}_{\sigma_j}(\mathbf{x}_P^*)$	5.00, 0.06	0.37, 0.05	25.96, 1.64
$d_{\mu_j}(\hat{y}_{\mu_j}(\mathbf{x}_P^*))$, $d_{\sigma_j}(\hat{y}_{\sigma_j}(\mathbf{x}_P^*))$	0.50, 0.36	0.45, 0.50	0.73, 0.36

† The $\hat{y}_{\sigma_j}$ and $d_{\sigma_j}(\hat{y}_{\sigma_j})$ values for the standard deviation responses are computed *a posteriori* at the given $\boldsymbol{x}_{DS}^*$, and are written in italic.

Proposed Approach : General Properties

- Advantages
  - Good Balance among Responses on Both Location and Dispersion Effects
  - Robust to Potential Dependencies among Responses
  - Physical Interpretation of λ
- Disadvantages
  - Unreasonable Solutions Possible

    e.g. Let $\boldsymbol{d} = (d_{\mu 1}, d_{\mu 2}, d_{\sigma 1}, d_{\sigma 2})$

    $\boldsymbol{d_1} = (0.5, 0.5, 0.5, 0.5)$ vs. $\boldsymbol{d_2} = (0.99, 0.99, 0.99, 0.49)$

    $\boldsymbol{d_1} = (0.5, 0.5, 0.5, 0.5)$ vs. $\boldsymbol{d_3} = (0.99, 0.99, 0.99, 0.50)$
  - Costs for Required Replication

Micro-Array Design (SS Young)

- How do you pick the sequences that you are going to use.
  - They have to be unique vs other sub-sequences and they have to work at common temperature/chemical fluid concentrations.
- Placement of sub sequences on the chip and the number of rep spots.
- A no-no is possible reuse of chips. How to wash would be a good DOE problem.
  - Technically I think the chips can be reused, but contract requires only one use.
- There is very large variation in response among the sub-sequences within a gene. Does this relate to the sequence in some way?
  - If so, it would be of interest to try to figure out the factors influencing the among sub-sequence variation. So the question would be DOE on the selection of the subsequence and then DOE on the assay conditions.

Where have all the Data gone?

- No need for data (Theoretical Development)
- Survey Sampling and Design of Experiment (Physical data collection)
- Computer Simulation (Experiment)
  - Statistical Simulation
    (Random Number generation)
  - Engineering Simulation
- Data from Internet
  - On-line auction
  - Search Engine

STILL QUESTION?

Send $500 to

- Dennis Lin

  University Distinguished Professor

  483 Business Building

  Department of Supply Chain & Information Systems

  Penn State University

- +1 814 865-0377 (phone)
- +1 814 863-7076 (fax)
- DKL5@psu.edu

(Customer Satisfaction or your money back!)