Response Surface Methodology: 50 Years Later

Dennis K.J. Lin
The Pennsylvania State University
DennisLin@psu.edu

Workshop on SASA Conference
30 November, 2006

Response Surface Methodology

- RSM Strategy
- Variable Screening Process
- Analysis of Response Surface
- Dual Response Surface and Multiple response Problems
- Others?

Response Surface Methodology

Box and Wilson (1951)

"On the Experimental Attainment of Optimum Condition," *JRSS-B*, **13**, 1-45.

Box and Hunter (1957)

"Multi-Factor Experimental Designs for Exploring Response Surface," *Annals of Mathematical Statistics*, **28**, 195-241.

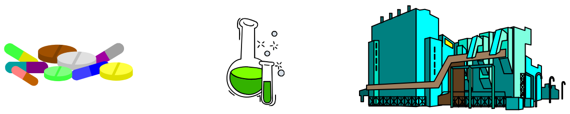
Ambitious Goal

- What is Response Surface Methodology?
- What type of problems they had in mind back to 1950?
- What was available in 1950?
- What type of problems today

  (50 years later)?
- What is available today?
- Can we do something significantly different?

What is RSM All About?

The Experimenter is like a person attempting to map the depth of the sea by making soundings at a limited number of places



Basic Approach

- If we are far away from the top, all we need is to find the direction for improvement...in this case, a first-order approximation may be sufficient.
- If we are close to the top, all we need is to find the exact location of the top...in this case, a more complicated model (such as a second-order model) is needed.

- If not much is known, you are likely to be in preliminary stage where first-order and screening designs give big benefits.

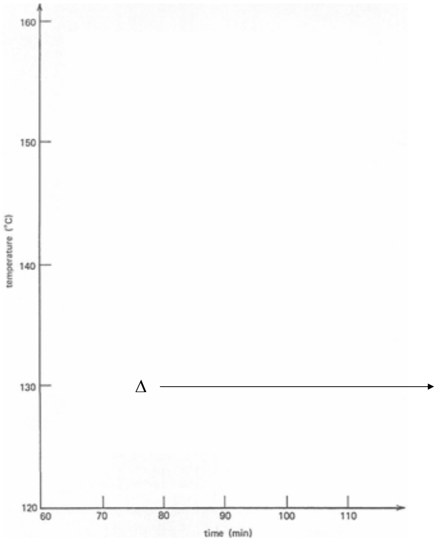
  When you are a long way from the top of the mountain, a plane may be a good approximation and you can probably use 1st-order linear approximation.

- If a lot if known, you may find more detailed study is necessary, in particular, careful study of maxima.

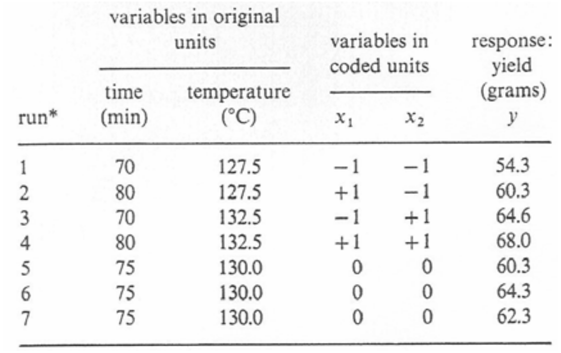
Illustrative Example (BH^2)

- Response (y): Yield
- Input Variable (x_1): time
- Input Variable (x_2): temperature

$$y=f(x_1, x_2)+\varepsilon$$



Current Setting:
x_1=75 min
x_2= 130 °C
➔ y=62.5



	variables in original units		variables in coded units		response: yield (grams)
run*	time (min)	temperature (°C)	x_1	x_2	y
1	70	127.5	−1	−1	54.3
2	80	127.5	+1	−1	60.3
3	70	132.5	−1	+1	64.6
4	80	132.5	+1	+1	68.0
5	75	130.0	0	0	60.3
6	75	130.0	0	0	64.3
7	75	130.0	0	0	62.3

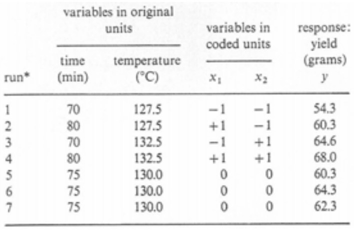
Fit

$$\hat{y} = \underset{(\pm 0.57)}{62.01} + \underset{(\pm 0.75)}{2.35x_1} + \underset{(\pm 0.75)}{4.5x_2}$$

$b_{12} = -0.65$ Nonsign.

$$b_{11} + b_{22} (= \bar{y}_f - \bar{y}_c)$$

= -0.50 Nonsign.



	variables in original units		variables in coded units		response: yield (grams)
run*	time (min)	temperature (°C)	x_1	x_2	y
1	70	127.5	−1	−1	54.3
2	80	127.5	+1	−1	60.3
3	70	132.5	−1	+1	64.6
4	80	132.5	+1	+1	68.0
5	75	130.0	0	0	60.3
6	75	130.0	0	0	64.3
7	75	130.0	0	0	62.3

Least Square Fitting:

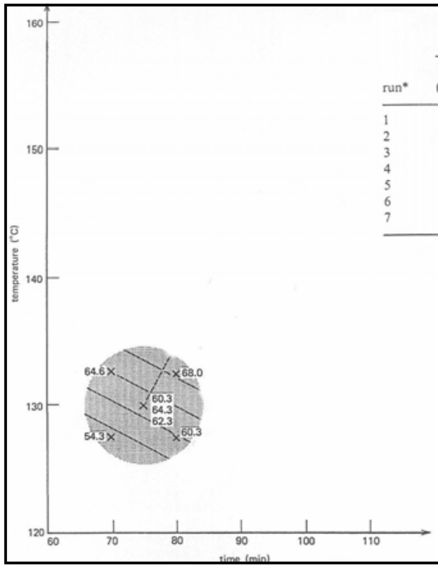
$y = 62.01+2.35x_1+4.50x_2 + \varepsilon$
$b_{12} = -0.65$ ($\pm$0.75)
$b_{11}+b_{22} = -0.50$ ($\pm$1.15)

Conclusion:

First-order model is adequate.

Action Taken:

Steepest Ascent
—direction for improvement
— 2.35:4.50 (or 1:1.91)



run*	variables in original units time (min)	temperature (°C)	variables in coded units x_1	x_2	response: yield (grams) y
1	70	127.5	−1	−1	54.3
2	80	127.5	+1	−1	60.3
3	70	132.5	−1	+1	64.6
4	80	132.5	+1	+1	68.0
5	75	130.0	0	0	60.3
6	75	130.0	0	0	64.3
7	75	130.0	0	0	62.3

Least Square Fitting:

$y = 62.01+2.35x_1+4.50x_2 + e$

$b_{12} = -0.65\ (\underline{+}0.75)$

$b_{11}+b_{22} = -0.50\ (\underline{+}1.15)$

Conclusion:

First-order model is adequate.

Action Taken:

Steepest Ascent

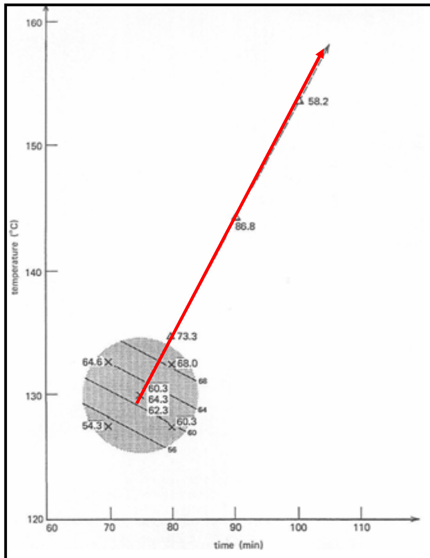
—direction for improvement

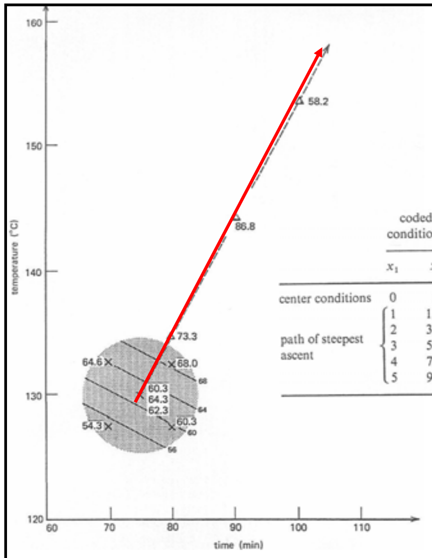
— 2.35:4.50 (or 1:1.91)

Run more experiments…

…following the direction of 1:1.91 (= 2.35 : 4.50)

	coded conditions x_1	x_2	time (min) t	temperature (°C) T	run	observed yield
center conditions	0	0	75	130.0	5, 6, 7	62.3 (average)
path of steepest ascent	1	1.91	80	134.8	8	73.3
	2	3.83	85	139.6		
	3	5.74	90	144.4	10	86.8
	4	7.66	95	149.1		
	5	9.57	100	153.9	9	58.2





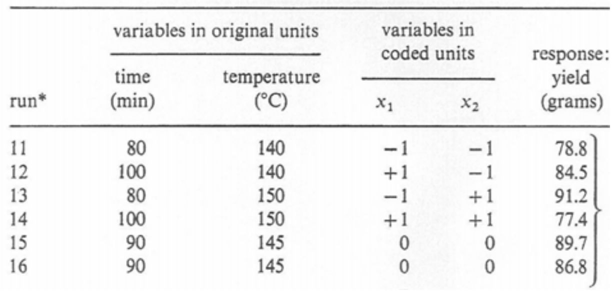
	coded conditions x_1	x_2	time (min) t	temperature (°C) T	run	observed yield
center conditions	0	0	75	130.0	5, 6, 7	62.3 (average)
path of steepest ascent	1	1.91	80	134.8	8	73.3
	2	3.83	85	139.6		
	3	5.74	90	144.4	10	86.8
	4	7.66	95	149.1		
	5	9.57	100	153.9	9	58.2

Conclusion:

Use Point #10 as
the new center point
and start all over again!!

Start all over again!!

--Use Point #10 as the new center point



	variables in original units		variables in coded units		
run*	time (min)	temperature (°C)	x_1	x_2	response: yield (grams)
11	80	140	−1	−1	78.8
12	100	140	+1	−1	84.5
13	80	150	−1	+1	91.2
14	100	150	+1	+1	77.4
15	90	145	0	0	89.7
16	90	145	0	0	86.8

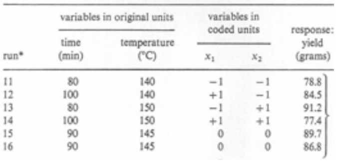
Fit

$$\hat{y} = \underset{(\pm 0.61)}{84.73} - \underset{(\pm 0.75)}{2.025}\, x_1 + \underset{(\pm 0.75)}{1.325}\, x_2$$

$$b_{12} = -4.88\ (\pm 0.75)$$

$$b_{11} + b_{22} = -5.28\ (\pm 1.15)$$

Significant!



	variables in original units		variables in coded units		
run*	time (min)	temperature (°C)	x_1	x_2	response: yield (grams)
11	80	140	−1	−1	78.8
12	100	140	+1	−1	84.5
13	80	150	−1	+1	91.2
14	100	150	+1	+1	77.4
15	90	145	0	0	89.7
16	90	145	0	0	86.8

Least Square Fitting:

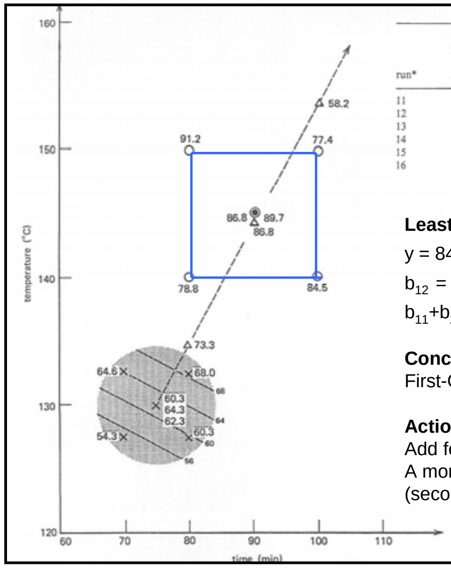
$y = 84.73 – 2.025x_1 + 1.325x_2 + \varepsilon$

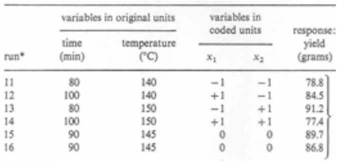
$b_{12} = –4.88\ (\underline{+}0.75)$

$b_{11}+b_{22} = –5.28\ (\underline{+}1.15)$

Conclusion:
First-Order Model is inadequate!

Action Taken:
Add few more points for fitting
A more complicated
(second-order) model.





	variables in original units		variables in coded units		
run*	time (min)	temperature (°C)	x_1	x_2	response: yield (grams)
11	80	140	−1	−1	78.8
12	100	140	+1	−1	84.5
13	80	150	−1	+1	91.2
14	100	150	+1	+1	77.4
15	90	145	0	0	89.7
16	90	145	0	0	86.8

Least Square Fitting:

$y = 84.73 – 2.025x_1 + 1.325x_2 + \varepsilon$

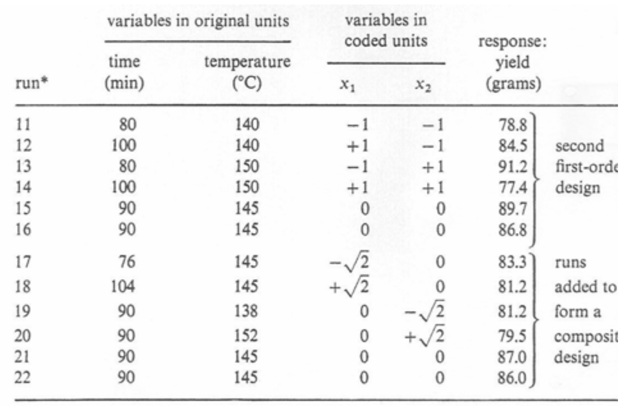
$b_{12} = –4.88\ (\underline{+}0.75)$

$b_{11}+b_{22} = –5.28\ (\underline{+}1.15)$

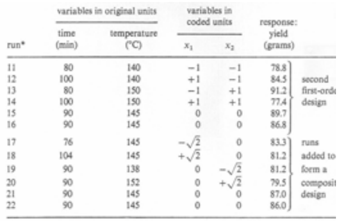
Conclusion:
First-Order Model is inadequate!

Action Taken:
Add few more points for fitting
A more complicated
(second-order) model.

Add few more points for fitting a second-order model.



	variables in original units		variables in coded units		response:	
run*	time (min)	temperature (°C)	x_1	x_2	yield (grams)	
11	80	140	-1	-1	78.8	
12	100	140	$+1$	-1	84.5	second
13	80	150	-1	$+1$	91.2	first-orde
14	100	150	$+1$	$+1$	77.4	design
15	90	145	0	0	89.7	
16	90	145	0	0	86.8	
17	76	145	$-\sqrt{2}$	0	83.3	runs
18	104	145	$+\sqrt{2}$	0	81.2	added to
19	90	138	0	$-\sqrt{2}$	81.2	form a
20	90	152	0	$+\sqrt{2}$	79.5	composit
21	90	145	0	0	87.0	design
22	90	145	0	0	86.0	



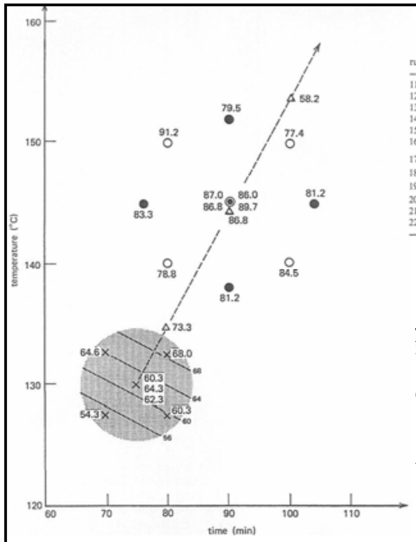
	variables in original units		variables in coded units		response:	
run*	time (min)	temperature (°C)	x_1	x_2	yield (grams)	
11	80	140	-1	-1	78.8	
12	100	140	$+1$	-1	84.5	second
13	80	150	-1	$+1$	91.2	first-orde
14	100	150	$+1$	$+1$	77.4	design
15	90	145	0	0	89.7	
16	90	145	0	0	86.8	
17	76	145	$-\sqrt{2}$	0	83.3	runs
18	104	145	$+\sqrt{2}$	0	81.2	added to
19	90	138	0	$-\sqrt{2}$	81.2	form a
20	90	152	0	$+\sqrt{2}$	79.5	composit
21	90	145	0	0	87.0	design
22	90	145	0	0	86.0	

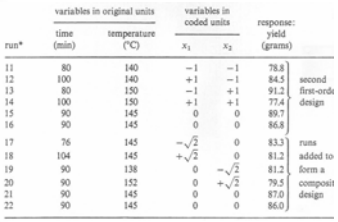
Least Square Fitting:

$$y = 87.36 - 1.39x_1 + 0.37x_2 + 2.15\,x_1^2 - 3.12\,x_2^2 - 4.88\,x_1\,x_2 + \varepsilon$$

Conclusion:
Second-Order model is adequate!

Action Taken:
Finding Optimal Setting!





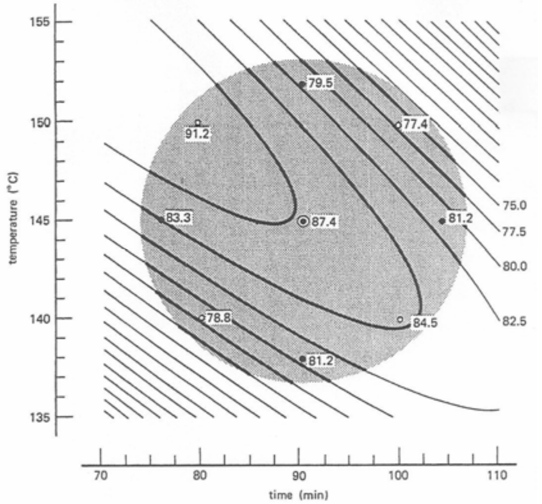
	variables in original units		variables in coded units		response:	
run*	time (min)	temperature (°C)	x_1	x_2	yield (grams)	
11	80	140	-1	-1	78.8	
12	100	140	$+1$	-1	84.5	second
13	80	150	-1	$+1$	91.2	first-orde
14	100	150	$+1$	$+1$	77.4	design
15	90	145	0	0	89.7	
16	90	145	0	0	86.8	
17	76	145	$-\sqrt{2}$	0	83.3	runs
18	104	145	$+\sqrt{2}$	0	81.2	added to
19	90	138	0	$-\sqrt{2}$	81.2	form a
20	90	152	0	$+\sqrt{2}$	79.5	composit
21	90	145	0	0	87.0	design
22	90	145	0	0	86.0	

Least Square Fitting:

$$y = 87.36 - 1.39x_1 + 0.37x_2 + 2.15\,x_1^2 - 3.12\,x_2^2 - 4.88\,x_1\,x_2 + \varepsilon$$

Conclusion:
Second-Order model is adequate!

Action Taken:
Finding Optimal Setting!



$$y = 87.36 - 1.39x_1 + 0.37x_2 + 2.15\,x_1^2 - 3.12\,x_2^2 - 4.88\,x_1\,x_2 + \varepsilon$$

Final Remarks

- The global optimum turns out to be

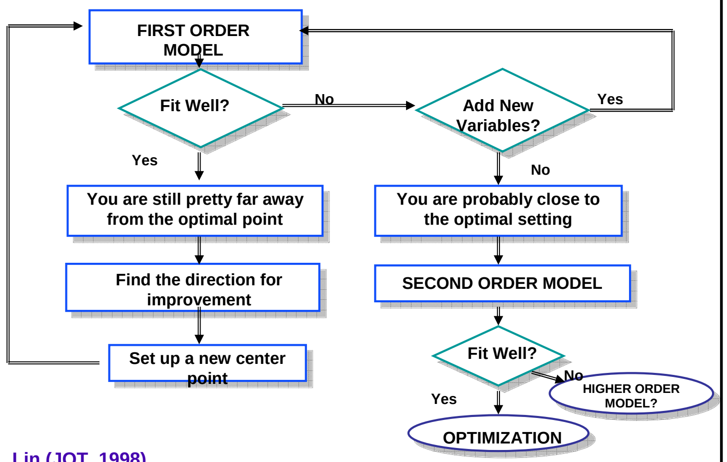
  x_1=80 minutes

  x_2=150 °C

  E(y)=91.2

  (as oppose to 62.5 at the beginning)

- Is such an optimal setting *feasible*?



FIRST ORDER MODEL

Fit Well?

No

Yes

Add New Variables?

Yes

No

You are still pretty far away from the optimal point

Find the direction for improvement

Set up a new center point

You are probably close to the optimal setting

SECOND ORDER MODEL

Fit Well?

No

HIGHER ORDER MODEL?

Yes

OPTIMIZATION

Lin (JQT, 1998)

Response Surface Methodology

(Box and Draper, 1987)

- WHICH (Screening)
- HOW (Empirical Model Building)
- WHY (Mechanistic Model Building)

RSM: General Steps

- Define
- Design
- Modeling
- Estimation
- Optimization
- Forecasting
- Confirmation

What are the issues?

- Data Collection:
  - What will be a good design? For what purpose?
- Data Analysis:
  - What will be a good model?
- Optimization:
  - Objective function?
- Confirmation

Response Surface Methodology Theoretical Formulation

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

$$x \in \Omega$$

- Objective
  - Find

    $x = x^*$ such that y is optimized.
- Basic Assumption/Belief
  - Life is Good

    — y is a smooth function of x

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

Issues to be Addressed

- $\boldsymbol{x}$: variable selection
  - Screening Input variables $x_1, x_2, \ldots, x_k$
- $\boldsymbol{f}$: model selection
- Θ: parameter estimation
- ε: error properties
- Ω: Experimental Region

Special Case-I

- ***x***: known
  - Input variables $x_1, x_2,\ldots, x_k$
- ***f***: model selection
  - First-Order Polynomial $y=\beta_0+\Sigma\beta_i x_i+\varepsilon$
- **Θ**: parameter estimation
  - Least square fitting
- ε: error properties
  - i.i.d. $N(0,\sigma^2)$
- Ω: Experimental Region
  - Correctly identified.

A (Typical) Special Case

- ***x***: known
  - Input variables $x_1, x_2,\ldots, x_k$
- ***f***: model selection
  - Second-Order Polynomial $y=\beta_0+\Sigma\beta_i x_i+\Sigma\beta_{ij} x_i x_j+\varepsilon$
- **Θ**: parameter estimation
  - Least square fitting
- ε: error properties
  - i.i.d. $N(0,\sigma^2)$
- Ω: Experimental Region
  - Correctly identified.

Devil's Advocate (Box, 1990)

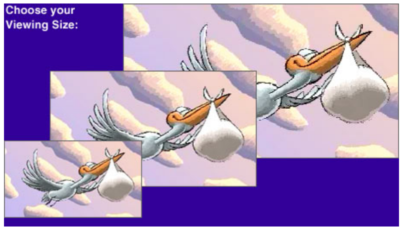
● One-at-a-time	● Large interaction
● Steepest ascent	● Many bumps
● Fractional factorial	● Large three-factor interaction
● Second-order fitting	● Exponential type response surface
● Grid mapping	● Flat plane with single point sticking out

Design: How to use the minimal cost to accomplish the experimental goal of

- Screening
- Empirical Model Building
- Linear model fitting in general
- Non-Linear model fitting
- Validation of *f* and θ
- Future trace
- All-In-One (copier/scanner/fax/printer/..)

Screening

Killing storks will not reduce the birth rate!



Design of Screening Experiments

- Two-Level Fractional Factorials
- Plackett & Burman Design (Hadamard Matrix)
- Two-Level Orthogonal Arrays
- Regular Simplex & T-optimal
- p-efficient Designs
- Supersaturated Designs

Lin(2003)

Before Experiment

$$y = f(x_1, \ldots, x_p, \underbrace{x_{p+1}, \ldots, x_k}) + \varepsilon$$

After Experiment

$$y = f(x_1, \ldots, x_p) + \varepsilon(x_{p+1}, \ldots, x_k)$$

$$p << k$$

MODEL

$$Y = \underset{\sim}{1} \cdot \mu + X \underset{\sim}{\beta} + \underset{\sim}{\varepsilon}$$

$Y_{n\times1}$: observable data
$X_{n\times k}$: design matrix
$\beta_{k\times1}$: parameter vector
$\varepsilon_{n\times1}$: noise

$N = \{i_1, i_2, \ldots, i_p\}$	inert factor
$A = \{i_{p+1}, i_{p+2}, \ldots, i_k\}$	active factor

$N \cup A = \{1, 2, \ldots, k\}$

Goal

Test $H_j : \beta_j = 0$ vs. $H_j^c : \beta_j \neq 0$

$$\begin{cases} H_j \text{ is true if } j \in N \\ H_j^c \text{ is true if } j \in A \end{cases}$$

How to Accomplish the Goal?

Criterion for Optimal Design?

Model/Analysis here:

First-Order Model?

ANOVA Model?

Examples

- ANONA Approach:
  - Orthogonal Array (n=4t)
- First-Order Model:
  - Minimal-Point Design (n=k+1)
- Significant Test Approach
  - Good estimate of σ!
- Others?

About Model Building

- Smoothing assumption in *f*.
- Typically polynomial model is assumed, as the empirical model building.
- Spline Fitting
- Artificial Neural Network
- Radial Basis Function
- Non- (Semi-) Parametric Fitting
- Optimality versus Robustness

Designs for Model Building

- Central Composite Design (CCD)
- Small Composite Design
- Box and Behnken Design
- Three-Level Design
- Uniform Design
- Others

$$y=\beta_0+\sum \beta_i x_i+\sum \beta_{ij} x_i x_j+\sum \beta_{ii} x_i^2+\varepsilon$$

Parameter Estimation

- Least Square Estimate
- Likelihood approach (with proper assumption on the distribution)
- Bayesian approach, when appropriate
- Black-Box approach, such as Artificial Neural Network

Assumption on Noise

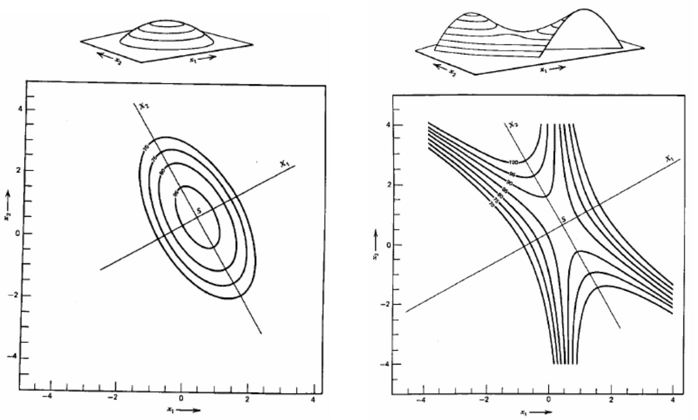
- i.i.d. $N(0, \sigma^2)$ Assumption
- Generalized Least square
- Generalized Linear model

- Bayesian Approach

Analysis of Response Surface

Objectives

- Overall Surface structure
- Optimal value of y
- Corresponding setup x*
- Future exploration



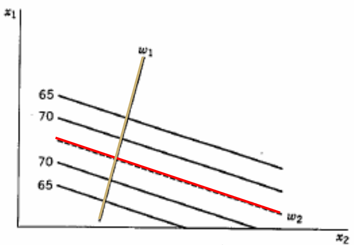
How About

- Goodness/Badness of fit
- Optimal y outside the current domain
- Confidence Region of y*
- Confidence Region of x*

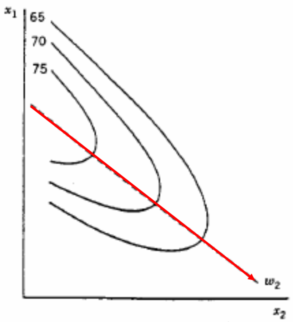
Second-Order Polynomial Model

- Estimation: β vs $\hat{\beta}$
- Bias: f vs $\hat{f}$
- Prediction: y_{max} vs $\hat{y}_{\max}$
- Prediction: x^* vs $\hat{x}^*$
- Point Estimate & Confidence Region (Sweet Spot)
- General f?

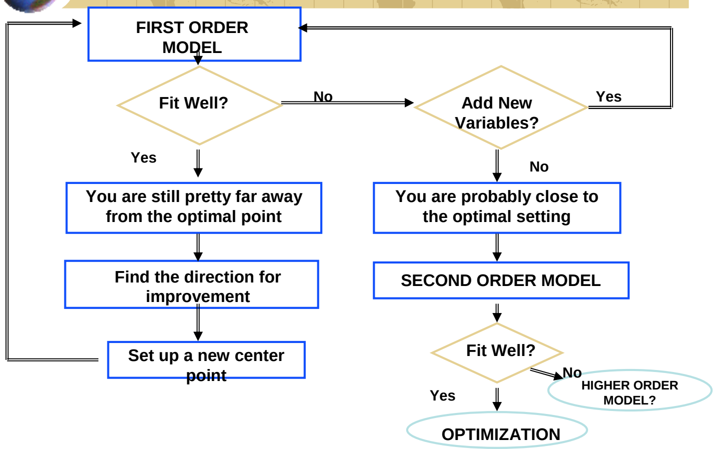
Ridge Systems



First-Order Surface



Second-Order Surface



FIRST ORDER MODEL

Fit Well?

No

Add New Variables?

Yes

Yes

You are still pretty far away from the optimal point

Find the direction for improvement

Set up a new center point

No

You are probably close to the optimal setting

SECOND ORDER MODEL

Fit Well?

No

HIGHER ORDER MODEL?

Yes

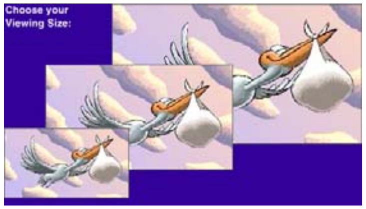
OPTIMIZATION

Lin (JQT, 1998)

Lecture 2:
Screening Experimentation

Screening

Killing storks will not reduce the birth rate!



Before Experiment

$$y = f(x_1,\dots,x_p,\underbrace{x_{p+1},\dots,x_k}) + \varepsilon$$

After Experiment

$$y = f(x_1,\dots,x_p) + \varepsilon(x_{p+1},\dots,x_k)$$

$$p << k$$

MODEL

$$Y = \underset{\sim}{1} \cdot \mu + X \underset{\sim}{\beta} + \underset{\sim}{\varepsilon}$$

$Y_{n\times 1}$: observable data
$X_{n\times k}$: design matrix
$\beta_{k\times 1}$: parameter vector
$\varepsilon_{n\times 1}$: noise

LET

$N = \{i_1, i_2, \dots, i_p\}$ inert factor
$A = \{i_{p+1}, i_{p+2}, \dots, i_k\}$ active factor
$N \cup A = \{1,2,\dots,k\}$

Goal

Test $H_j : \beta_j = 0 \ \text{vs.}\ H_j^c : \beta_j \neq 0$

$$\begin{cases} H_j \text{ is true if } j \in N \\ H_j^c \text{ is true if } j \in A \end{cases}$$

Design of Screening Experiments

- Two-Level Fractional Factorials
- Plackett & Burman Design
  (Hadamard Matrix)
- Two-Level Orthogonal Arrays
- Regular Simplex & T-optimal
- p-efficient Designs
- Supersaturated Designs

Lin(2003)

Full Factorial Design

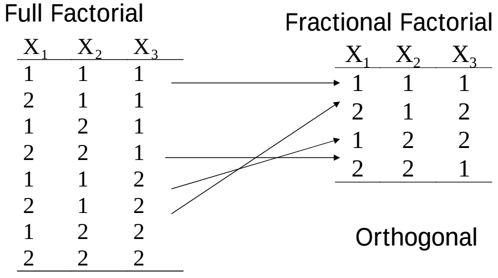
Full Factorial Design
(All possible Combinations)
Multi-Factors
Example: X_1 has two possibilities (1 or 2)
X_2 has two possibilities (1or 2)
X_3 has two possibilities (1or 2)

X_1	X_2	X_3
1	1	1
2	1	1
1	2	1
2	2	1
1	1	2
2	1	2
1	2	2
2	2	2

A Total of 2^3 experimental runs

Fractional Factorial Designs

Ex: X_1 is two-level (1or 2)
X_2 is two-level (1or 2)
X_3 is two level (1or 2)



Full Factorial

X_1	X_2	X_3
1	1	1
2	1	1
1	2	1
2	2	1
1	1	2
2	1	2
1	2	2
2	2	2

Fractional Factorial

X_1	X_2	X_3
1	1	1
2	1	2
1	2	2
2	2	1

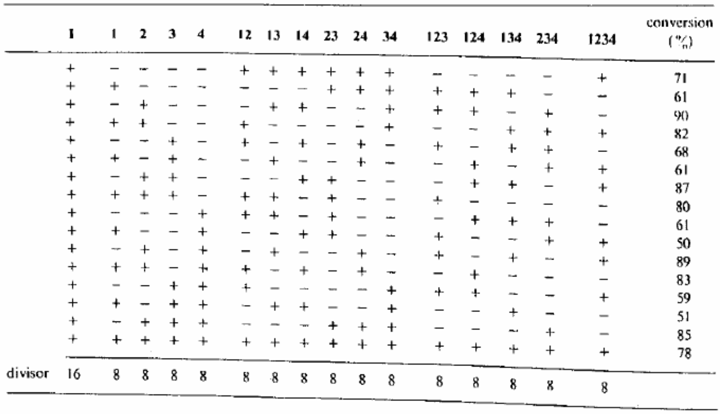
Orthogonal

❖ R.A. Fisher (1920)
❖ F. Yates

Issues to be Considered

- Why Two-Level?
- Why Full Factorial (Advantages)?
- Why NOT Full Factorial (Disadvantage)?
- Addendum?

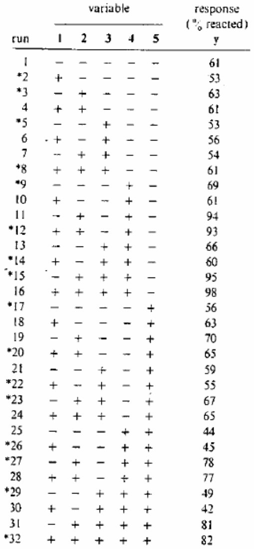
Example of Full Factorial Design



I	1	2	3	4	12	13	14	23	24	34	123	124	134	234	1234	conversion (%)
+	−	−	−	−	+	+	+	+	+	+	−	−	−	−	+	71
+	+	−	−	−	−	−	−	+	+	+	+	+	+	−	−	61
+	−	+	−	−	−	+	+	−	−	+	+	+	−	+	−	90
+	+	+	−	−	+	−	−	−	−	+	−	−	+	+	+	82
+	−	−	+	−	+	−	+	−	+	−	+	−	+	+	−	68
+	+	−	+	−	−	+	−	−	+	−	−	+	−	+	+	61
+	−	+	+	−	−	−	+	+	−	−	−	+	+	−	+	87
+	+	+	+	−	+	+	−	+	−	−	+	−	−	−	−	80
+	−	−	−	+	+	+	−	+	−	−	−	+	+	+	−	61
+	+	−	−	+	−	−	+	+	−	−	+	−	−	+	+	50
+	−	+	−	+	−	+	−	−	+	−	+	−	+	−	+	89
+	+	+	−	+	+	−	+	−	+	−	−	+	−	−	−	83
+	−	−	+	+	+	−	−	−	−	+	+	+	−	−	+	59
+	+	−	+	+	−	+	+	−	−	+	−	−	+	−	−	51
+	−	+	+	+	−	−	−	+	+	+	−	−	−	+	−	85
+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	78
divisor 16	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8	

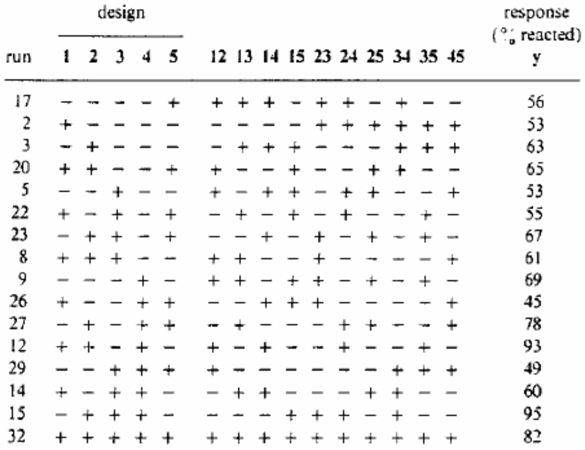
A Second Example

- **x_1**: Feed Rate
  (10 & 15 liters/min)
- **x_2**: Catalyst
  (1% & 2%)
- **x_3**: Agitation rate
  (100rpm & 120rpm)
- **x_4**: Temperature
  (140°C & 180°C)
- **x_5**: Concentration
  (3% & 6%)



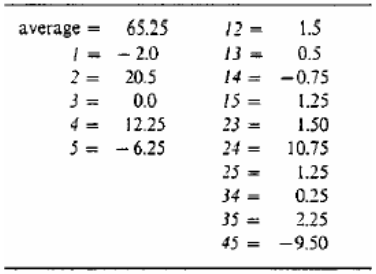
run	variable 1	2	3	4	5	response (% reacted) y
1	−	−	−	−	−	61
*2	+	−	−	−	−	53
*3	−	+	−	−	−	63
4	+	+	−	−	−	61
*5	−	−	+	−	−	53
6	+	−	+	−	−	56
7	−	+	+	−	−	54
*8	+	+	+	−	−	61
*9	−	−	−	+	−	69
10	+	−	−	+	−	61
11	−	+	−	+	−	94
*12	+	+	−	+	−	93
13	−	−	+	+	−	66
*14	+	−	+	+	−	60
*15	−	+	+	+	−	95
16	+	+	+	+	−	98
*17	−	−	−	−	+	56
18	+	−	−	−	+	63
19	−	+	−	−	+	70
*20	+	+	−	−	+	65
21	−	−	+	−	+	59
*22	+	−	+	−	+	55
*23	−	+	+	−	+	67
24	+	+	+	−	+	65
25	−	−	−	+	+	44
*26	+	−	−	+	+	45
*27	−	+	−	+	+	78
28	+	+	−	+	+	77
*29	−	−	+	+	+	49
30	+	−	+	+	+	42
31	−	+	+	+	+	81
*32	+	+	+	+	+	82

Half-Fraction



run	design 1	2	3	4	5	12	13	14	15	23	24	25	34	35	45	response (% reacted) y
17	−	−	−	−	+	+	+	+	−	+	+	−	+	−	−	56
2	+	−	−	−	−	−	−	−	−	+	+	+	+	+	+	53
3	−	+	−	−	−	−	+	+	+	−	−	−	+	+	+	63
20	+	+	−	−	+	+	−	−	+	−	−	+	+	−	−	65
5	−	−	+	−	−	+	−	+	+	−	+	+	−	−	+	53
22	+	−	+	−	+	−	+	−	+	−	+	−	−	+	−	55
23	−	+	+	−	+	−	−	+	−	+	−	+	−	+	−	67
8	+	+	+	−	−	+	+	−	−	+	−	−	−	−	+	61
9	−	−	−	+	−	+	+	−	+	+	−	+	−	+	−	69
26	+	−	−	+	+	−	−	+	+	+	−	−	−	−	+	45
27	−	+	−	+	+	−	+	−	−	−	+	+	−	−	+	78
12	+	+	−	+	−	+	−	+	−	−	+	−	−	+	−	93
29	−	−	+	+	+	+	−	−	−	−	−	−	+	+	+	49
14	+	−	+	+	−	−	+	+	−	−	−	+	+	−	−	60
15	−	+	+	+	−	−	−	−	+	+	+	−	+	−	−	95
32	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	82

- What to Gain?
- What to Lose?
- Will this design serve for your purpose?
- Which half-fraction?



average = 65.25
1 = −2.0
2 = 20.5
3 = 0.0
4 = 12.25
5 = −6.25

12 = 1.5
13 = 0.5
14 = −0.75
15 = 1.25
23 = 1.50
24 = 10.75
25 = 1.25
34 = 0.25
35 = 2.25
45 = −9.50

What are the issues here?

- Which fraction?
- What to compare (criteria)?
- Regular Orthogonal Fractions
- Non-regular Orthogonal Fractions
- Irregular (non-orthogonal) Fractions
  - Addelman (1961), *Technometrics*, 479-496.

Illustration 2-level (2^{6-2}) designs

D=

1	2	3	4	5=1234	6=124
-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	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	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

I = 12345
= 1246
= 356

Word Length Pattern
WLP=(0,0,1,1,1,0)

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)

Word Length Pattern

- Given a regular two-level fractional factorial design D
- Word Length Pattern

  $W(D)=(A_1(D), A_2(D),\ldots, A_k(D))$

  where $A_i(D)$ is the number of words in the defining relation whose length is *i*.

Resolution

- Resolution III Design
  - Main-Effect Design
  - Minimal-Point Design
- Resolution IV Design
  - Webb, Cheng.
- Resolution V Designs
- Relationship among different resolution designs
- Isomorphism

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.

Optimality Criteria

- Estimability and Efficiency
- Resolution
- Aberration
- Generalized Aberration
- Clear Factor
- Projectivity
- Model Robustness
- Estimation Capacity

Irregular Fractional Factorials

- What does Orthogonality really for???
- Three-Quarter Fraction
- Addelman Design
  - "Irregular fractions of the 2^n factorial experiments," *Technometrics*, **3**, 479-496.
- Minimal-Point Design
- p-efficient Design
- Cyclic Design

More about
Resolution
Aberration
Generalized Aberration

Plackett & Burman Designs

$$\mathbf{PB}_{12} = \begin{bmatrix} + & + & - & + & + & + & - & - & - & + & - \\ + & - & + & + & + & - & - & - & + & - & + \\ - & + & + & + & - & - & - & + & - & + & + \\ + & + & + & - & - & - & + & - & + & + & - \\ + & + & - & - & - & + & - & + & + & - & + \\ + & - & - & - & + & - & + & + & - & + & + \\ - & - & - & + & - & + & + & - & + & + & + \\ - & - & + & - & + & + & - & + & + & + & - \\ - & + & - & + & + & - & + & + & + & - & - \\ + & - & + & + & - & + & + & + & - & - & - \\ - & + & + & - & + & + & + & - & - & - & + \\ - & - & - & - & - & - & - & - & - & - & - \end{bmatrix}$$

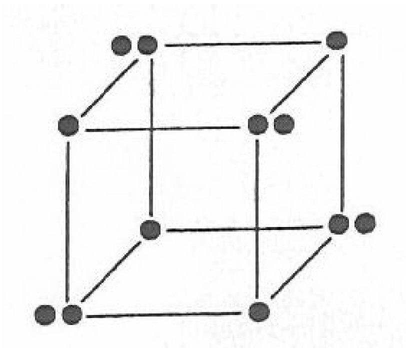
Hadamard Matrices

- Given a n by n, two-symbol (+-1) square matrix, what is the largest determinant possible???
- A matrix H such that H'H=nI.
- n must be a multiple of four, except n=1 and 2.

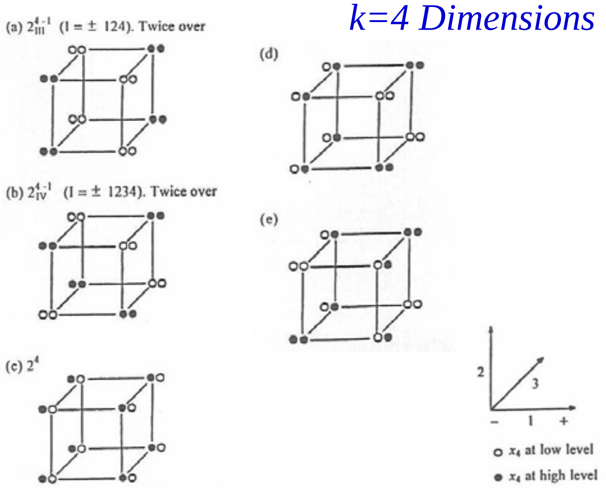
Some Research on Hadamard Matrix

- Construction (n=268?)
- Non-equivalent Hadamard Matrices

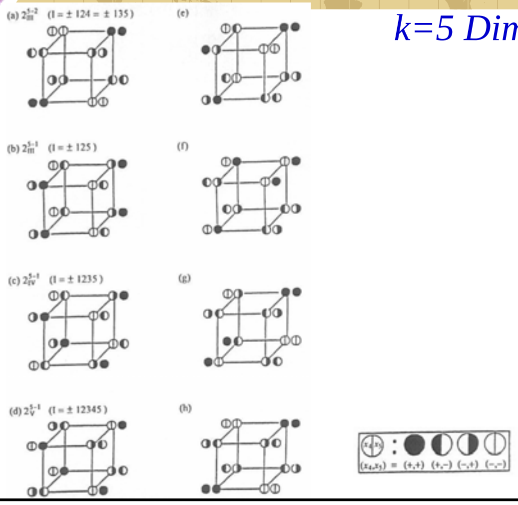
The 12-run Plackett & Burman into k=3 Dimensions



The 12-run Plackett & Burman into k=4 Dimensions



The 12-run Plackett & Burman into k=5 Dimensions



Projection Properties

q	Design	Description
2	2.1	$2^2 \times 3$ (2^2 design with 3 replicates)
3	3.1	$2^3 + \frac{1}{2}2^3$ (2^3 design plus 2^{3-1} design)
4	4.1	Add one more runs to form a 2_{IV}^{4-1} design
		Add five more runs to form a 2^4 design
5	5.1	Add two more runs to form a 2_{III}^{5-2} design
		Add six more runs to form a 2_{V}^{5-1} design
	5.2	Add two more runs to form a 2_{III}^{5-2} design
		Add eight more runs to form a 2_{IV}^{5-1} design
		Add ten more runs to form a 2_{V}^{5-1} design

Regular Simplex Design

u	x_{1u}	x_{2u}	$\cdots$	x_{ku}
1	−1	−1	$\cdots$	−1
2	1	−1	$\cdots$	−1
3	0	2	$\cdots$	−1
4	0	0	$\cdots$	−1
.	.	.	$\cdots$	.
.	.	.	$\cdots$	.
n	0	0	$\cdots$	k

Regular Simplex Design *(k=3)*

$$\begin{pmatrix} -1 & -1 & -1 \\ 1 & -1 & -1 \\ 0 & 2 & -1 \\ 0 & 0 & 3 \end{pmatrix} \qquad \begin{pmatrix} -\sqrt{2} & -\sqrt{2/3} & -1/\sqrt{3} \\ \sqrt{2} & -\sqrt{2/3} & -1/\sqrt{3} \\ 0 & 2\sqrt{2/3} & -1/\sqrt{3} \\ 0 & 0 & \sqrt{3} \end{pmatrix}$$

Regular Simplex Design

Pros and Cons

Optimal Designs

D-optimality: maximize $\| \mathbf{X'X} \| = \lambda_1 \times \lambda_2 \times \cdots \times \lambda_k$.

A-optimality: minimize $\text{trace}(\mathbf{X'X})^{-1} = \sum_{i=1}^{k} \lambda_i^{-1}$.

E-optimality: maximize the smallest eigenvalue of the $\mathbf{X'X}$ matrix.

G-optimality: minimize the maximum prediction variance over the operation region.

V-optimality: minimize the average prediction variance over the operation region.

T-optimal Design

$$\mathbf{X} = \begin{bmatrix} x_1 & x_2 & x_3 & \cdots & x_k \\ x_2 & x_3 & x_4 & \cdots & x_1 \\ x_3 & x_4 & x_5 & \cdots & x_2 \\ \cdot & \cdot & \cdot & \cdots & \cdot \\ \cdot & \cdot & \cdot & \cdots & \cdot \\ x_k & x_1 & x_2 & \cdots & x_{k-1} \\ -1 & -1 & -1 & \cdots & -1 \end{bmatrix}$$

Optimality Criterion?

$X_{(k+1)} - X_{(1)}$

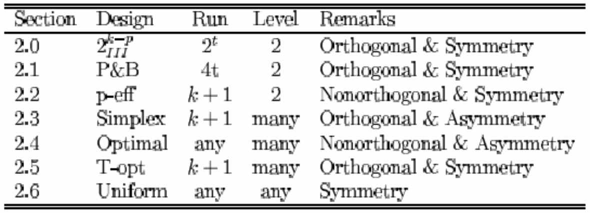
T-optimal Design (k=6)

$$\begin{pmatrix} -1.06 & 0.61 & -1.06 & -0.08 & 1.30 & 1.30 \\ 0.61 & -1.06 & -0.08 & 1.30 & 1.30 & -1.06 \\ -1.06 & -0.08 & 1.30 & 1.30 & -1.06 & 0.61 \\ -0.08 & 1.30 & 1.30 & -1.06 & 0.61 & -1.06 \\ 1.30 & 1.30 & -1.06 & 0.61 & -1.06 & -0.08 \\ 1.30 & -1.06 & 0.61 & -1.06 & -0.08 & 1.30 \\ -1 & -1 & -1 & -1 & -1 & -1 \end{pmatrix}$$

T-optimal Design

k	$(x_1, x_2, \cdots, x_k)$
1	(1.00)
2	(1.37, -0.37)
3	(-1.00, 1.00, 1.00)
4	(0.81, -1.43, 0.81, 0.81)
5	(-0.79, 0.20, -1.00, 1.29, 1.29)
6	(-1.06, 0.61, -1.06, -0.08, 1.30. 1.30)
7	(1.00, -1.00, -1.00, 1.00, 1.00, 1.00, -1.00)
8	(-1.01, 0.40, -1.01, 0.44, -0.58, -0.43, 1.60, 1.60)
9	(-1.12, -0.07, -1.12, -0.07, 1.24, -1.12, 0.77, 1.24, 1.24)

Summary



Section	Design	Run	Level	Remarks
2.0	2^{k-p}_{III}	2^t	2	Orthogonal & Symmetry
2.1	P&B	4t	2	Orthogonal & Symmetry
2.2	p-eff	$k+1$	2	Nonorthogonal & Symmetry
2.3	Simplex	$k+1$	many	Orthogonal & Asymmetry
2.4	Optimal	any	many	Nonorthogonal & Asymmetry
2.5	T-opt	$k+1$	many	Orthogonal & Symmetry
2.6	Uniform	any	any	Symmetry

Research Potentials

- Objective of Screening Design
- Minimal Effort (experimental runs)
- High Efficient (optimal criterion)
- Geometry Property (orthogonality & projection)

Supersaturated Design

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 (*Technometrics, 1993*)

From Saturated to Supersaturated

No degree of freedom for σ
to
Negative degree of freedom for σ

Some Basic Approaches (Design Construction)

- Orthogonal Array-Based
- Group Screening
- Non-Orthogonal Array-Based
- Combinatorial Approach
- Optimization Approach

Recent Applications in SSD

- (Nano-) Manufacturing
- Computer Experiments
- Numerical Analysis
- e-Business
- Marketing Survey
- High Dimensional Integration

Current (Factor) Screening Procedure

Professional Knowledge

THERE ARE ALWAYS MORE VARIABLES THAN WE CAN HANDLE !!!

$$\underbrace{x_1, x_2, \ldots, x_k}_{\text{experimental variables}} \qquad \underbrace{x_{k+1}, \ldots, x_m}_{\text{how to handle these variables?}}$$

- Ignore
- Fix
- Randomization, Blocking, ...
- Supersaturated Design

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*

Supersaturated Design —how to study *k* parameters with *n(<<k)* observations?

- *What for ?*
- *How to construct ?*
- *How to analyze ?*
- *Limitations ?*
- *Does it really work ?*

A New Class of Supersaturated Designs

Dennis K. J. Lin
Department of Statistics
The University of Tennessee
Knoxville, TN 37996

Supersaturated designs are useful in situations in which the number of active factors is very small compared to the total number of factors being considered. In this article, a new class of supersaturated designs is constructed using half fractions of Hadamard matrices. When a Hadamard matrix of order *N* is used, such a design can investigate up to *N - 2* factors in *N/2* runs. Results are given for $N \leq 60$. Extension to larger *N* is straightforward. These designs are superior to other existing supersaturated designs and are easy to construct. An example with real data is used to illustrate the ideas.

KEY WORDS: Hadamard matrices; Plackett and Burman designs; Random balance designs.

UTK Technical Report, 1991

Life After Screening

- Follow-Up Experiment
- Projection Properties

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	+	-	-	-	-	-	-	-	-	-	-	-

Half Fraction of H_{12}

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

$n = N/2 = 6$ $k = N\text{-}2 = 10$

Supersaturated Design Example

Half Fraction of William's (1968) Data

Run	Factor																							y
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*)

$$H_{n\times n} = \begin{bmatrix} 1 & 1 & H_1 \\ 1 & -1 & H_2 \end{bmatrix}_{n\times n}$$

H_1 and H_2 are isormophic?
$(n - 2) \times n/2$

Lin (1991)

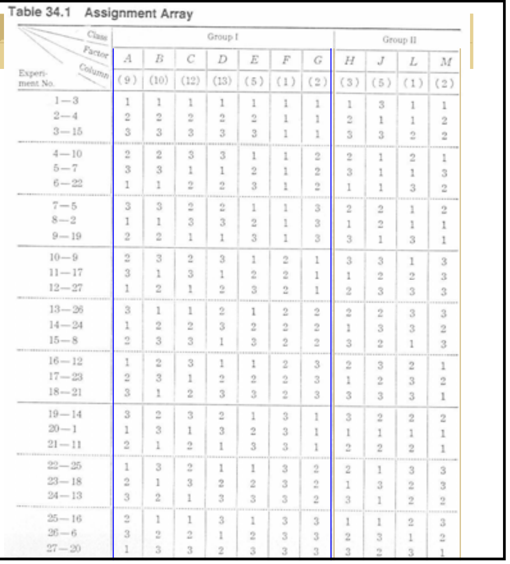
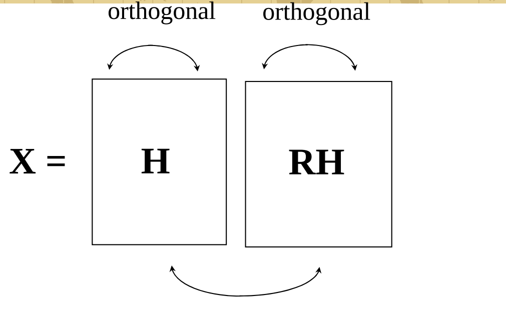


Table 34.1 Assignment Array

Class / Factor / Column / Experiment No.	Group I							Group II			
	A	*B*	*C*	*D*	*E*	*F*	*G*	*H*	*J*	*L*	*M*
	(9)	(10)	(12)	(13)	(5)	(1)	(2)	(3)	(5)	(1)	(2)
1—3	1	1	1	1	1	1	1	1	3	1	1
2—4	2	2	2	2	2	1	1	2	1	1	2
3—15	3	3	3	3	3	1	1	3	3	2	2
4—10	2	2	3	3	1	1	2	2	1	2	1
5—7	3	3	1	1	2	1	2	3	1	1	3
6—22	1	1	2	2	3	1	2	1	1	3	2
7—5	3	3	2	2	1	1	3	2	2	1	2
8—2	1	1	3	3	2	1	3	1	2	1	1
9—19	2	2	1	1	3	1	3	3	1	3	1
10—9	2	3	2	3	1	2	1	3	3	1	3
11—17	3	1	3	1	2	2	1	1	2	2	3
12—27	1	2	1	2	3	2	1	2	3	3	3
13—26	3	1	1	2	1	2	2	2	2	3	3
14—24	1	2	2	3	2	2	2	1	3	3	2
15—8	2	3	3	1	3	2	2	3	2	1	3
16—12	1	2	3	1	1	2	3	2	3	2	1
17—23	2	3	1	2	2	2	3	1	2	3	2
18—21	3	1	2	3	3	2	3	3	3	3	1
19—14	3	2	3	2	1	3	1	3	2	2	2
20—1	1	3	1	3	2	3	1	1	1	1	1
21—11	2	1	2	1	3	3	1	2	2	2	1
22—25	1	3	2	1	1	3	2	2	1	3	3
23—18	2	1	3	2	2	3	2	1	3	2	3
24—13	3	2	1	3	3	3	2	3	1	2	2
25—16	2	1	1	3	1	3	3	1	1	2	3
26—6	3	2	2	1	2	3	3	2	3	1	2
27—20	1	3	3	2	3	3	3	3	2	3	1

Random
Balance
Design

Taguchi (1986)



not orthogonal
Thus Permute rows of RH
to minimize $E(s^2)$, say.

Supersaturated Design including an orthogonal base.

$X = \begin{bmatrix} H & RHC \end{bmatrix}$

↑ matrix for column selection to get rid of fully aliased columns

EXAMPLES:

(1) $R = \underset{\text{product}}{D}(h_i)$ Wu (1993)

(2) $R = \underset{\text{permute}}{P}$ Tang & Wu (1993)

(3) $R = PD(h_i)$

(4) $R = \frac{1}{n} H \underset{\uparrow}{a} H'$

nonequivalent Hadamard mx

Design Criteria

What is a "good" supersaturated design?

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-criterion
- Deng, Lin and Wang (1994): resolution rank
- Balkin and Lin (1997):
  Graphical Comparison (Harmonic mean of eigens)
- Fang, Lin and Liu (2002):

CRITERIA FOR SUPERSATURATED DESIGNS

(C1) s = max s_{ij}

(C2) $E(s^2) = \sum s_{ij}^2 / \binom{k}{2}$

(C3) $\rho = \sum r_{ij}^2 / \binom{k}{2}$

(C4) D-criterion $= \frac{1}{\binom{k}{c}} \sum \det(X_s' X_s)^{-1}$

(C5) A-criterion $= \frac{1}{\binom{k}{c}} \sum \operatorname{trace}(X_s' X_s)^{-1}$

(C6) E-criterion $= \frac{1}{\binom{k}{c}} \sum \lambda_{(c)}(X_s' X_s)^{-1}$

(C7) B-criterion $= \frac{1}{\binom{k}{c}} \sum \beta_{s-i}'(X_{s-i}' X_{s-i})^g \beta_{s-i}$

(C8) resolution rank.

Deng& Lin (1994)

Recent Design Criteria

Supersaturated Design

- Uniformity
- Generalized Minimum Aberration
- Majorization
- E(f_{NOD})
- Projection Properties (D_f, A_f, G_f, etc)
- E(χ^2), as an extension of E(s^2)
- Minimax s_{ij}
- Model Robustness

Recent Design Criteria

Supersaturated Design

- Minimum Moment Aberration
- E(d_2)
- G_2-Aberration
- Asymptotic Power Properties
- Orthogonal-Based
- Factor-Covering
- Marginally Over-saturated
- Balance matrix

Life After Screening: Projection Properties

Resolution rank (r-rank)

Example: E(f_{NOD})

$$E(f_{NOD}) = \sum_{1 \le i < j \le m} f_{NOD}^{ij} \Big/ \binom{m}{2}$$

where $$f_{NOD}^{ij} = \sum_{u=1}^{q_i} \sum_{v=1}^{q_j} \left(n_{uv}^{(ij)} - \frac{n}{q_i q_j} \right)^2$$

Lower Bound of E(f_{NOD})

Theorem 1. *For any design* $X \in \mathscr{U}(n; q_1, \ldots, q_m)$,

$$\begin{aligned} E(f_{NOD}) &= \frac{\sum_{k,l=1,k \neq l}^{n} \lambda_{kl}^2}{m(m-1)} + C(n, q_1, \ldots, q_m) \\ &\geq \frac{n(\sum_{j=1}^{m} n/q_j - m)^2}{m(m-1)(n-1)} + C(n, q_1, \ldots, q_m), \end{aligned}$$

where $C(n, q_1, \ldots, q_m) = \frac{nm}{m-1} - \frac{1}{m(m-1)} \left(\sum_{i=1}^{m} \frac{n^2}{q_i} + \sum_{i,j=1,j \neq i}^{m} \frac{n^2}{q_i q_j} \right)$

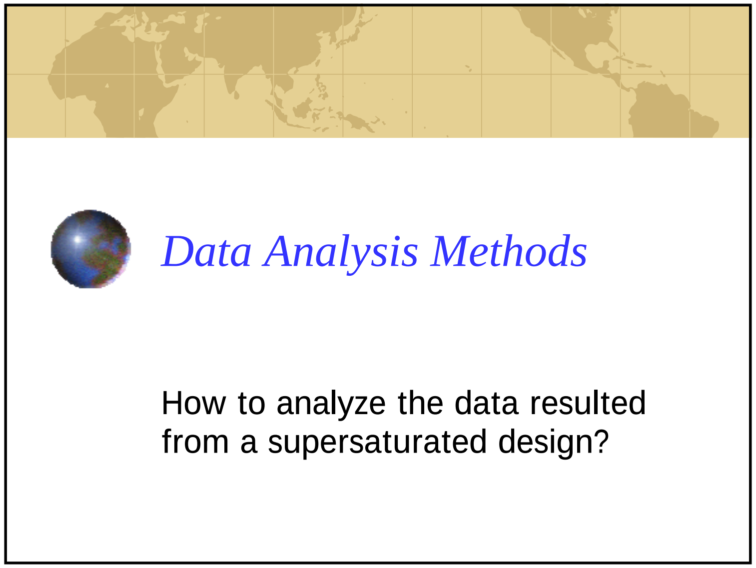
Connection with Previous Criteria

Corollary 1. *For any design* $X \in \mathscr{U}(n; q^m)$,

$$E(f_{NOD}) \geq \frac{mn}{(m-1)(n-1)} \left(\frac{n}{q} - 1 \right)^2 + \frac{n}{m-1} \left(m - \frac{n}{q} \right) - \left(\frac{n}{q} \right)^2,$$

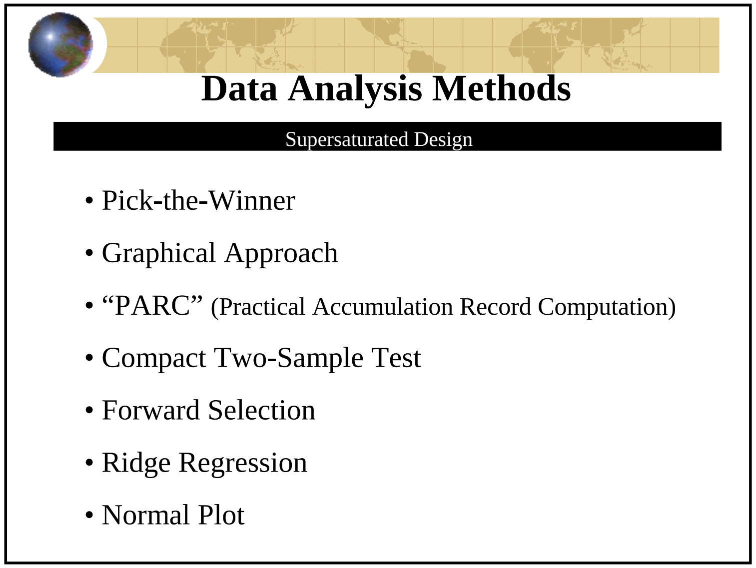
$$E(f_{NOD}) = \frac{n}{9} \text{ave } \chi^2, \ \textit{when } q_i = 3$$

$$E(f_{NOD}) = \frac{1}{4} E(s^2), \ \textit{when } q_i = 2,$$



Data Analysis Methods

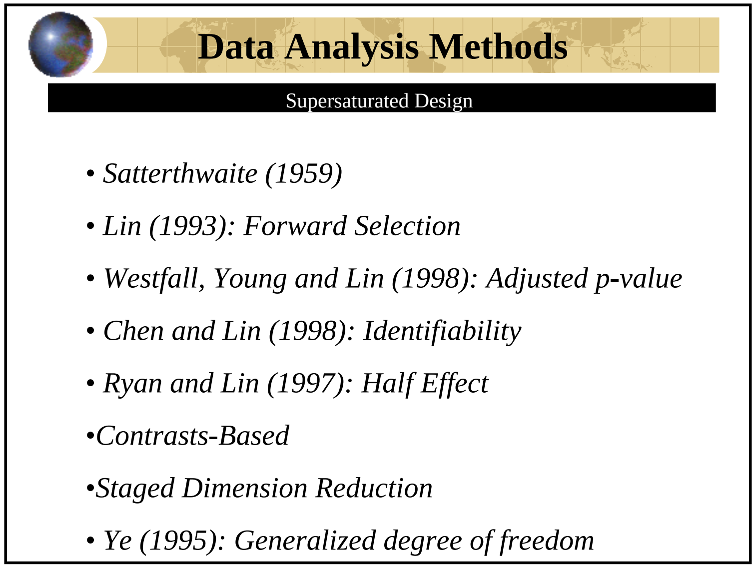
How to analyze the data resulted from a supersaturated design?



Data Analysis Methods

Supersaturated Design

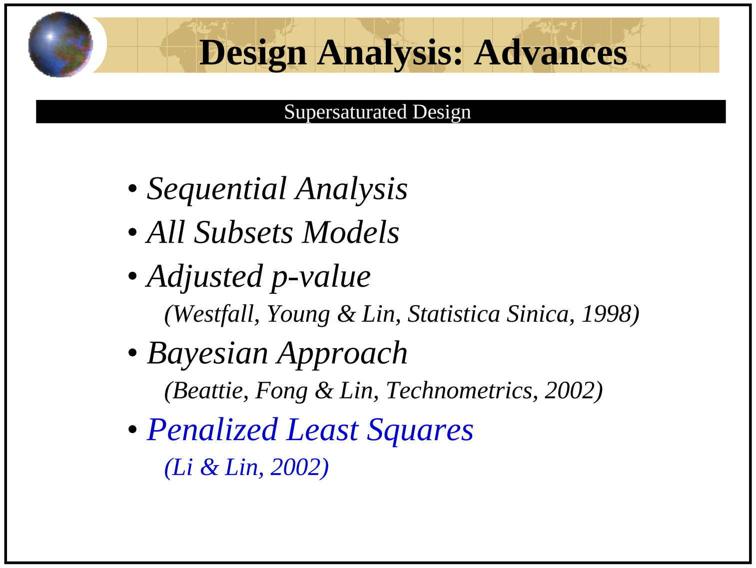
- Pick-the-Winner
- Graphical Approach
- "PARC" (Practical Accumulation Record Computation)
- Compact Two-Sample Test
- Forward Selection
- Ridge Regression
- Normal Plot



Data Analysis Methods

Supersaturated Design

- *Satterthwaite (1959)*
- *Lin (1993): Forward Selection*
- *Westfall, Young and Lin (1998): Adjusted p-value*
- *Chen and Lin (1998): Identifiability*
- *Ryan and Lin (1997): Half Effect*
- *Contrasts-Based*
- *Staged Dimension Reduction*
- *Ye (1995): Generalized degree of freedom*



Design Analysis: Advances

Supersaturated Design

- *Sequential Analysis*
- *All Subsets Models*
- *Adjusted p-value*
  (Westfall, Young & Lin, Statistica Sinica, 1998)
- *Bayesian Approach*
  (Beattie, Fong & Lin, Technometrics, 2002)
- *Penalized Least Squares*
  (Li & Lin, 2002)

MODEL

$$Y = \underset{\sim}{1} \cdot \mu + X \underset{\sim}{\beta} + \underset{\sim}{\varepsilon}$$

$Y_{n\times 1}$: observable data
$X_{n\times k}$: design matrix
$\beta_{k\times 1}$: parameter vector
$\varepsilon_{n\times 1}$: noise

LET

$N = \{i_1, i_2, \ldots, i_p\}$ inert factor
$A = \{i_{p+1}, i_{p+2}, \ldots, i_k\}$ active factor
$N \cup A = \{1, 2, \ldots, k\}$

Goal

Test $H_j : \beta_j = 0$ vs. $H_j^c : \beta_j \neq 0$

$$\begin{cases} H_j \text{ is true if } j \in N \\ H_j^c \text{ is true if } j \in A \end{cases}$$

Application to F.S. in S.S. Design

Method: Perform F.S. (ordinary) to determine entry sequence: F-values are $F_1^{(1)}, F_2^{(2)}, \ldots,$ corresponding to factors $X_{i1}, X_{i2}, \ldots$.

Multiple Testing Algorithm:

i. Compute resampling distribution of max $F_j^{*(1)}$ for first variable entered under complete null. *P*-value for first variable entered is $P(\max F_j^{*(1)} \geq F_1^{(1)})$.

ii. Compute resampling distribution of max $F_j^{*(2)}, j \in \{1, \ldots, k\} - i_1$, assuming all effects null but X_{i_1}. (Force X_{i_1} into all fits.) *P*-value for second variable entered is $P(\max_{j \in \{1,\ldots k\} - i_1} F_j^{*(2)} \geq F_2^{(2)})$.

iii. Compute resampling distribution of max $F_j^{*(3)}, j \in \{1, \ldots, k\} - \{i_1, i_2\}$, assuming all effects null but X_{i_1}, X_{i2}. (Force X_{i_1} and X_{i2} into all fits.) *P*-value for third variable entered is $P(\max_{j \in \{1,\ldots k\} - \{i_1, i_2\}} F_j^{*(3)} \geq F_3^{(3)})$.

NOTES ON ALGORITHM (Adjusted p-value)

- Attributable to Forsythe *et al.* and Miller.
- Re-sampling may be parametric (sample from normal distribution) or nonparametric (bootstrap sampling of residuals).
- Parametric re-sampling: Generate *Y* ~ N(0, I)*. At step *l*, compute $\max_{j \in \{1,\ldots k\} - \{i_1,\ldots,i_{l-1}\}} F_j^{*(l)}$.

  Compare to original $F_l^{(l)}$.
- The method is conditional on original order of entry.
- The re-sampling *p*-value are *forced* to be monotonic: e.g., if $\tilde{p}_2 < \tilde{p}_1$, define $\tilde{p}_2 = \tilde{p}_1$. Benefit: Protection of FWE under complete null.

Penalized Least Squares *(Li and Lin, 2003)*

Model

$$Y_i = \mathbf{x}_i^T \beta + \varepsilon_i$$

Penalized Likelihood (Fan and Li, 1999)

$$\frac{1}{n} \sum_{i=1}^{n} \log f(y_i, \mathbf{x}_i^T \beta) - \sum_{j=1}^{d} p_\lambda(|\beta_j|)$$

Becomes

$$Q(\beta) \equiv \frac{1}{2n} \sum_{i=1}^{n} (Y_i - \mathbf{x}_i^T \beta)^2 + \sum_{j=1}^{d} p_\lambda(|\beta_j|$$

Algorithmic

Supersaturated Design

Lin (1991, 1995): Pair-wise Optimality

Nguyen (1996): Exchange Algorithm

Li and Wu (1997): Column-wise and Pair-wise Algorithm

Church (1993): Projection Properties

Jones (2000): JMP Product

Spotlight Interaction Effects in Main Effect Plans: A Supersaturated Design Approach

Dennis K.J. Lin
Department of Statistics
University of Tennessee
Knoxville, TN 37996-0532

ABSTRACT

In a traditional screening experiment, a first-order model is commonly assumed; i.e., all interaction effects are tentatively ignored. The construction of first-order main-effect designs that are optimal in some sense has received a great deal of attention in the literature. However, the conventional wisdom on such a main-effect design can be misleading, if any interaction effect is presented. With no additional experimental cost, this paper shows how to spotlight interaction effects in these so-called "main-effect" designs. It is shown that the proposed method is superior to other existing approaches. Comparisons are made with an example for illustration. Limitations and further research directions are also discussed.

Key words: Effect Sparsity; Normal Plot; Plackett and Burman designs; Screening; Stepwise Regression.

- Hamada and Wu (1992)
- Hen and Wang (1994)
- Wu (1993)
- Westfall, Young and Lin(1995)

- **Estimated Main Effects**

A	0.3258		
B	0.2938		
C	-0.2458		
D	-0.5162	8	0.4458
E	0.1498	9	0.4525
F	0.9152	10	0.0805
G	0.1832	11	-0.2422

Main Effect Model:

$$(1)\ \hat{y} = \underset{(0.1616)}{5.73} + \underset{(0.1616)}{0.458}\,F$$

$R^2 = 44.5\%$
$S = .5596$
$R_a^2 = 39\%$

$$(2)\ \hat{y} = \underset{(0.1470)}{5.73} - \underset{(0.1470)}{0.258}\,D + \underset{(0.1470)}{0.458}\,F$$

$R^2 = 58.7\%$
$S = .5091$
$R_a^2 = 49.5\%$

• Estimated Interaction Effects

AB	.5578	BD	-.2375	CG	.3881
AC	-.5078	BE	-.0782	DE	-.0215
AD	-.1315	BF	-.0555	DF	-.0882
AE	-.9075	BG	-.2075	DG	.4838
AF	-.0515	CD	-.5152	EF	-.1735
AG	-.2575	CE	.1042	EG	-.1715
BC	-.5838	CF	.1282	FG	-.9175

Main + Interaction Effect Model:

Hamada & Wu (1992)

$$(1)\ \hat{y} = 5.73 + 0.458\,F - \underset{(0.0750)}{0.459}\,FG$$

$R^2 = 89.3\%$
$S = 0.2596$

$$(2)\ \hat{y} = \underset{(0.0726)}{5.73} + \underset{(0.0726)}{0.458}\,F + \underset{\substack{(p=0.243)\\ NS}}{0.0916}\,G - \underset{(0.0726)}{0.459}\,FG$$

$R^2 = 91\%$
$S = 0.2515$

$$(3)\ \hat{y} = \underset{(0.05)}{5.73} - \underset{\substack{(0.055)\\ [P-value=0.209]\\ NS}}{0.0761}\,D + \underset{(0.053)}{0.401}\,F - \underset{(0.055)}{0.377}\,FG - \underset{(0.059)}{0.169}\,AE$$

$R^2 = 96.3\%$
$S = 0.1732$

$$(4)\ \hat{y} = \underset{(0.0528)}{5.73} + \underset{(0.056)}{0.394}\,F - \underset{(0.056)}{0.395}\,FG - \underset{(0.060)}{0.191}\,AE$$

$R^2 = 95.3\%$
$S = 0.1828$

Example: William's Data

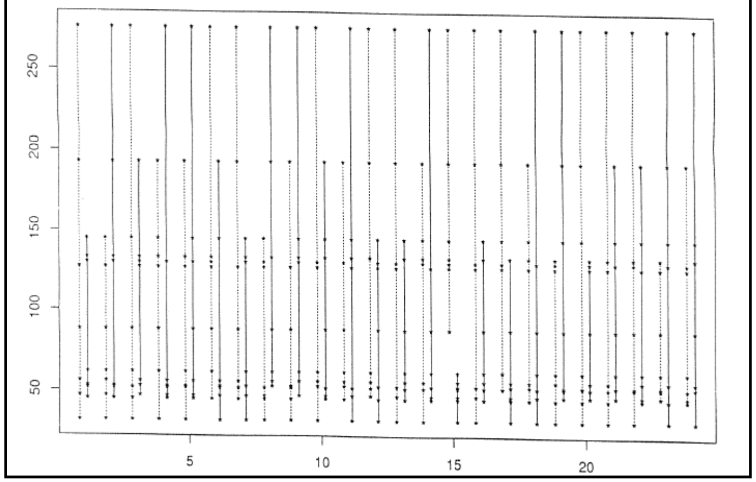
Supersaturated Design Example

Half Fraction of William's (1968) Data

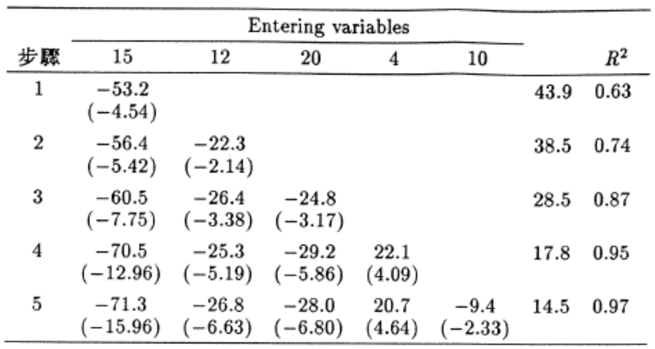
	Factor																								
Run	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	y
1	+	+	+	-	-	-	+	+	+	+	+	-	+	-	-	+	+	-	-	+	-	-	-	+	133
2	+	-	-	-	-	-	+	+	+	-	-	-	+	+	+	+	-	+	-	-	+	+	-	-	62
3	+	+	-	+	+	-	-	-	-	+	-	+	+	+	+	+	+	-	-	-	-	+	+	-	45
4	+	+	-	+	-	+	-	-	-	+	+	-	+	-	+	+	-	+	+	+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

Lin (*Technometrics, 1993*)

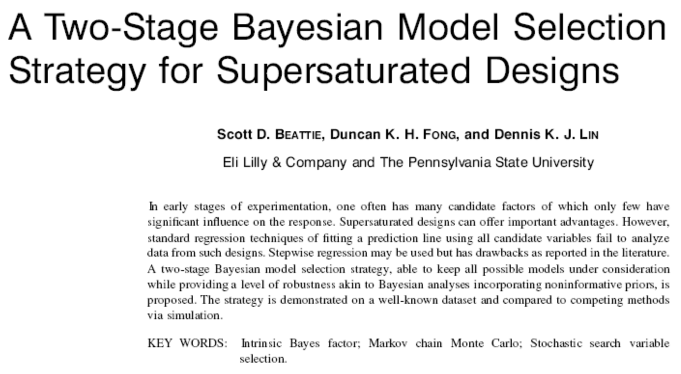
Graphical Approach



Stepwise Regression



	Entering variables						
步驟	15	12	20	4	10		R^2
1	−53.2 (−4.54)					43.9	0.63
2	−56.4 (−5.42)	−22.3 (−2.14)				38.5	0.74
3	−60.5 (−7.75)	−26.4 (−3.38)	−24.8 (−3.17)			28.5	0.87
4	−70.5 (−12.96)	−25.3 (−5.19)	−29.2 (−5.86)	22.1 (4.09)		17.8	0.95
5	−71.3 (−15.96)	−26.8 (−6.63)	−28.0 (−6.80)	20.7 (4.64)	−9.4 (−2.33)	14.5	0.97



A Two-Stage Bayesian Model Selection Strategy for Supersaturated Designs

Scott D. Beattie, Duncan K. H. Fong, and Dennis K. J. Lin

Eli Lilly & Company and The Pennsylvania State University

In early stages of experimentation, one often has many candidate factors of which only few have significant influence on the response. Supersaturated designs can offer important advantages. However, standard regression techniques of fitting a prediction line using all candidate variables fail to analyze data from such designs. Stepwise regression may be used but has drawbacks as reported in the literature. A two-stage Bayesian model selection strategy, able to keep all possible models under consideration while providing a level of robustness akin to Bayesian analyses incorporating noninformative priors, is proposed. The strategy is demonstrated on a well-known dataset and compared to competing methods via simulation.

KEY WORDS: Intrinsic Bayes factor; Markov chain Monte Carlo; Stochastic search variable selection.

SSVS & IBF

Table 4. Comparative Results for Analyses of the Williams Half-Fraction

Model Selection Method	*Factors Identified as Important*
Williams final model (Williams 1968)	4, 10, 14, 19
Forward selection w/modified *p* (Westfall et al. 1998)	14
Forward selection w/$p = .05$	14
Stepwise w/$p = .05$	14
Stepwise using AIC	14
SSVS ($\tau_i = \hat{\sigma}_{\beta_i}$, $c_i = 5$)	14
SSVS ($\tau_i = \hat{\sigma}_{\beta_i}$, $c_i = 10$)	14
SSVS ($\tau_i = .10\hat{\sigma}_{\beta_i}$, $c_i = 100$)	4, 12, 14, 19
SSVS ($\tau_i = .10\hat{\sigma}_{\beta_i}$, $c_i = 500$)	14
IBF using 4, 12, 14, 19 from SSVS	4, 12, 14, 19

Penalized Least Squares

$$Q(\boldsymbol{\beta}) \equiv \frac{1}{2n}\sum_{i=1}^{n}(Y_i - \mathbf{x}_i^T\boldsymbol{\beta})^2 + \sum_{j=1}^{d} p_\lambda(|\beta_j|)$$

- SCAD (*Smoothly Clipped Absolute Deviation*, Fan, 1997)

$$p'_\lambda(\beta) = \lambda\left\{I(\beta \le \lambda) + \frac{(a\lambda - \beta)_+}{(a-1)\lambda} I(\beta > \lambda)\right\}$$

- λ is to be estimated (e.g., via GCV of Wahba, 1977)

$$\mathrm{GCV}(\lambda) = \frac{1}{n}\frac{\|\mathbf{y} - \mathbf{X}\widehat{\boldsymbol{\beta}}(\lambda)\|^2}{\{1 - e(\lambda)/n\}^2}$$

$$\text{and } \widehat{\lambda} = \mathrm{argmin}_\lambda\{\mathrm{GCV}(\lambda)\}$$

Results via SCAD

Posterior Model Probabilities

Model	Prob.	R^2
4 12 15 20	0.0266	0.955
4 10 12 15 20	0.0259	0.973
4 10 11 12 15 20	0.0158	0.987
4 12 15 20 21	0.1200	0.969
4 11 12 15 20	0.0082	0.966

The Final model selected by SCAD *(λ=6.5673)*

Factor	Intercept	X4	X12	X15	X20
$\hat{\beta}$	102.7857	20.1084	-25.3946	-69.5738	-28.7967
$SE(\hat{\beta})$	4.5377	4.6965	4.6557	5.1075	4.6965

Comparisons via Simulation

Method	True Model Identified Rate	Smallest Effect Identified Rate	Avg. Size Median	Mean
Case I: One Active Effects				
SSVS(1/10,500)	40.5%	99%	2	3.1
SSVS(1/10,500)/IBF	61%	98%	1	2.5
SCAD	75.6%	100%	1	1.7
Case II: Three Active Effects				
SSVS(1/10, 500)	8.6%	30%	3	4.7
SSVS(1/10,500)/IBF	8.0%	28%	3	4.2
SCAD	74.7%	98.5%	3	3.3
Case III: Five Active Effects				
SSVS(1/10, 500)	36.4%	84%	6	8.0
SSVS(1/10,500)/IBF	40.7%	75%	5	5.6
SCAD	69.7%	99.4%	5	5.4

Other Construction Methods

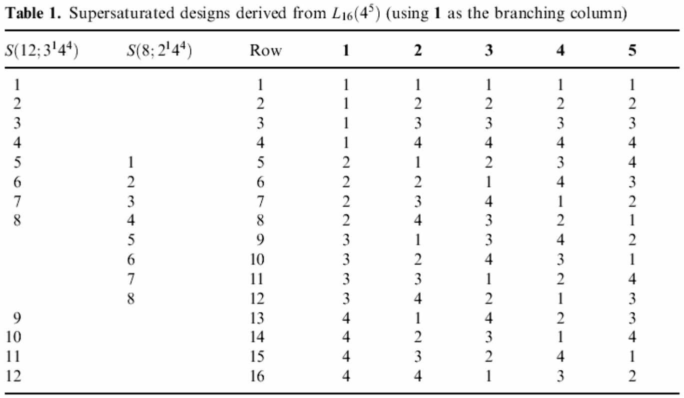
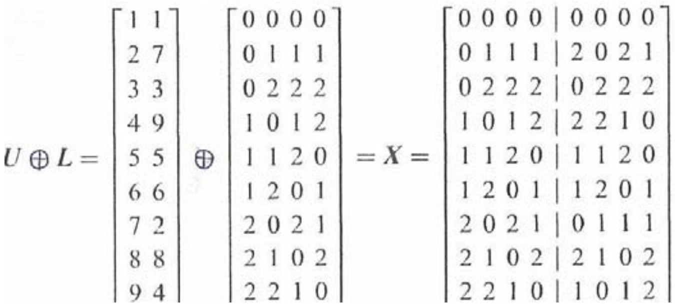


Table 1. Supersaturated designs derived from $L_{16}(4^5)$ (using **1** as the branching column)

$S(12;3^1 4^4)$	$S(8;2^1 4^4)$	Row	**1**	**2**	**3**	**4**	**5**
1		1	1	1	1	1	1
2		2	1	2	2	2	2
3		3	1	3	3	3	3
4		4	1	4	4	4	4
5	1	5	2	1	2	3	4
6	2	6	2	2	1	4	3
7	3	7	2	3	4	1	2
8	4	8	2	4	3	2	1
	5	9	3	1	3	4	2
	6	10	3	2	4	3	1
	7	11	3	3	1	2	4
	8	12	3	4	2	1	3
9		13	4	1	4	2	3
10		14	4	2	3	1	4
11		15	4	3	2	4	1
12		16	4	4	1	3	2

UD OD SSD



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

Fang, Lin & Ma (2000)

SSD: Looking Ahead

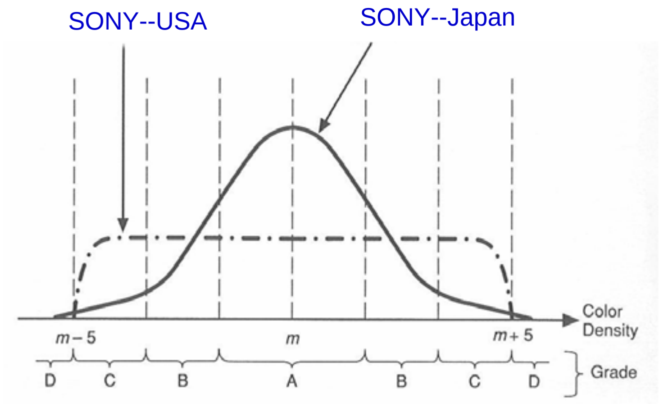
There will be more and more factors & parameters in the future experimental investigations!!!

New Methodology/Thinking is needed!

Lecture 3:
Taguchi Method &
Dual Response Analysis

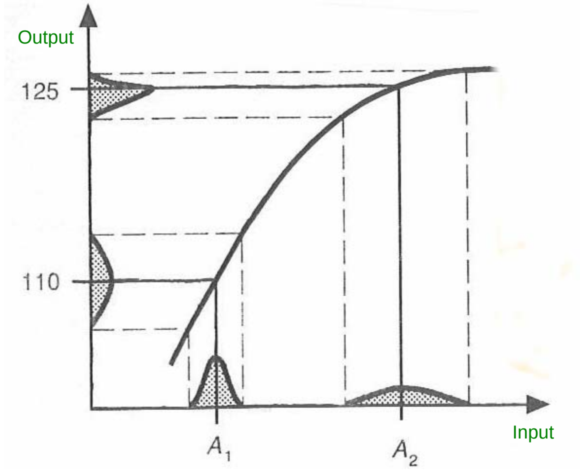
What is Quality?

- Fashion
- Reliability
- Yield (Productivity)
- Defective
- Capacity
- Optimization
- Variation Deduction
- Customer Satisfaction
- Market Share

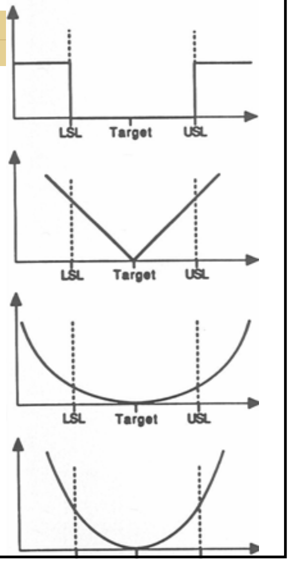


Taguchi Method

- Loss Function
- Robust Design
  (Inner/Outer Array Design)
- S/N Ratio



Quality Loss Function



Robust (Taguchi) Design

- What do the customers want?
- What the customers don’t want?
- Variation Deduction
  - Example: Sony
  - Example: Ina Tile

An overview of noise variables

- Controllable variables are process factors that can be controlled precisely.

- Noise variables are factors that cannot be controlled precisely. They introduce non-negligible error into the process response.

Examples of noise variables:

Experiment	Control variables	Noise variables
baking a cake	ingredients	oven temperature
tablet optimization	particle size	ambient humidity
poultry growth optimization	% protein, % carbohydrate	food intake

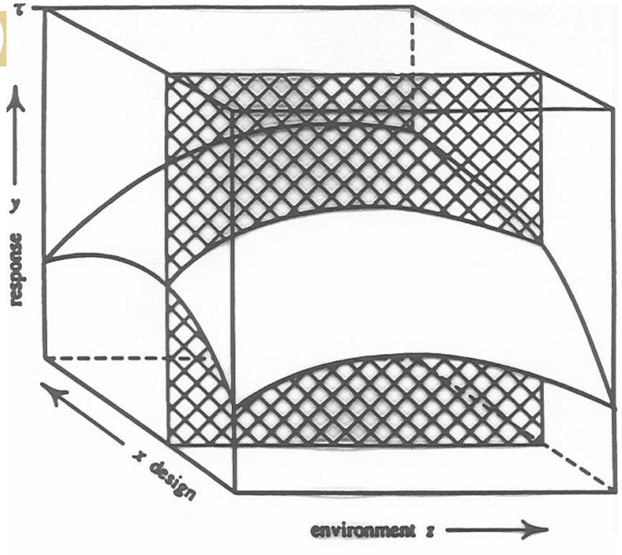
- In robust optimization we would like to configure the controllable factors such that ideally we have a process mean close to our target with low variance.

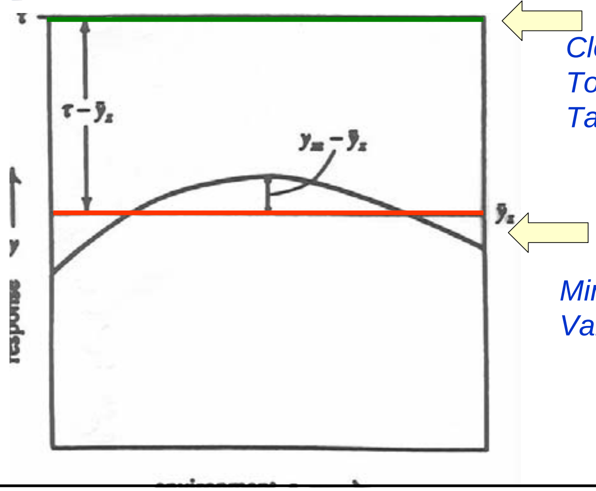
X_1	X_2	X_3
1	1	1
2	1	1
1	2	1
2	2	1
1	1	2
2	1	2
1	2	2
2	2	2

- One Observation
  Y_1
- Several Replicates
  $Y_{11}, Y_{12}, \ldots, Y_{1n}$
- Designing these Replicates

Z_1	Z_2	**Obs**
1	1	Y_{11}
2	1	Y_{12}
1	2	Y_{13}
2	2	Y_{14}

- Compound Orthogonal Arrays
- Uniform Design
  Wang, Fang and Lin (1995)





Close
To
Target

Minimum
Variation

Inner/Outer Array: Design, Analysis and Optimization

Printer Process Example

X_1	X_2	X_3	Y_1	Y_2	Y_3	X_1	X_2	X_3	Y_1	Y_2	Y_3	X_1	X_2	X_3	Y_1
-1	-1	-1	34	10	28	-1	-1	0	81	81	81	-1	-1	1	364
0	-1	-1	115	116	130	0	-1	0	90	122	93	0	-1	1	232
1	-1	-1	192	186	263	1	-1	0	319	376	376	1	-1	1	408
-1	0	-1	82	88	88	-1	0	0	180	180	154	-1	0	1	182
0	0	-1	44	178	188	0	0	0	372	372	372	0	0	1	507
1	0	-1	322	350	350	1	0	0	541	568	396	1	0	1	846
-1	1	-1	141	110	86	-1	1	0	288	192	312	-1	1	1	236
0	1	-1	259	251	259	0	1	0	432	336	513	0	1	1	660
1	1	-1	290	280	245	1	1	0	713	725	754	1	1	1	878

Myers, R.H. and Carter, W.H. (1973)
Response Surface Techniques
for Dual Response System
Technometrics, **15**, 301-317.

Primary Response	location
Secondary Response	dispersion

Printer Process Example

X_1	**X_2**	**X_3**	**Mean**	**sd**	**X_1**	**X_2**	**X_3**	**Mean**	**sd**	**X_1**	**X_2**	**X_3**	**Mean**	**sd**
-1	-1	-1	24	12.49	-1	-1	0	81	0	-1	-1	1	220.7	133.8
0	-1	-1	120.3	8.39	0	-1	0	101.7	17.67	0	-1	1	239.7	23.46
1	-1	-1	213.7	42.8	1	-1	0	357	32.91	1	-1	1	422	18.52
-1	0	-1	86	3.46	-1	0	0	171.3	15.01	-1	0	1	199	29.45
0	0	-1	136.7	80.41	0	0	0	372	0	0	0	1	485.3	44.64
1	0	-1	340	16.17	1	0	0	501.7	92.5	1	0	1	673.7	158.2
-1	1	-1	112.3	27.57	-1	1	0	264	63.5	-1	1	1	176.7	55.51
0	1	-1	256.3	4.62	0	1	0	427	88.61	0	1	1	501	138.9
1	1	-1	271.7	23.63	1	1	0	730.7	21.08	1	1	1	1010	142.5

$$\eta_p = \beta_0 + \sum_{i=1}^{k} \beta_i x_i + \sum_{i=1}^{k} \beta_{ii} x_i^2 + \sum_{i<j}^{k} \sum \beta_{ij} x_i x_j + \varepsilon_p$$

$$\eta_s = \gamma_0 + \sum_{i=1}^{k} \gamma_i x_i + \sum_{i=1}^{k} \gamma_{ii} x_i^2 + \sum_{i<j}^{k} \sum \gamma_{ij} x_i x_j + \varepsilon_s$$

$$\begin{aligned}\hat{\omega}_\mu = {} & 327.6 + 177.0x_1 + 109.4x_2 + 131.5x_3 + 32.0x_1^2 \\ & - 22.4x_2^2 - 29.1x_3^2 + 66.0x_1x_2 \\ & + 75.5x_1x_3 + 43.6x_2x_3.\end{aligned}$$

$$\begin{aligned}\hat{\omega}_\sigma = {} & 34.9 + 11.5x_1 + 15.3x_2 + 29.2x_3 + 4.2x_1^2 \\ & - 1.3x_2^2 + 16.8x_3^2 + 7.7x_1x_2 + 5.1x_1x_3 + 14.1x_2x_3.\end{aligned}$$

$$\min \quad \hat{\omega}_\sigma$$

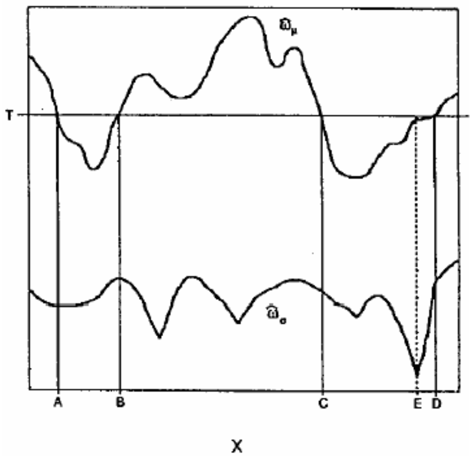
$$\text{subject to} \quad \hat{\omega}_\mu = \mu_0.$$

$$(x_1, x_2, x_3) = (0.614,\ 0.228,\ 0.1)$$

$$\hat{\omega}_\mu = 500$$

$$\hat{\omega}_\sigma = 51.77$$

An Illustrative Example



Step 1. *Find a model fitting for* ω_μ *and* ω_σ *(both are functions of x).*

Step 2. *Find x, such that*

$$MSE = (\hat{\omega}_\mu - T)^2 + \hat{\omega}_\sigma^2$$

is minimized.

Method	Optimal Setting	$\hat{\omega}_{\mu}$	$\hat{\omega}_{\sigma}^{2}$	MSE
Vining & Myers	(0.614, 0.228, 0.1)	500	2679.70	2679.70
MSE method	(1.0, 0.07, -0.25)	494.44	1974.02	2005.14

Best Subset Model:

$$\hat{\omega}_{\mu} = 314.667 + 177.0x_1 + 109.426x_2 + 131.463x_3 + 66.028x_1x_2 + 75.472x_1x_3 + 43.583x_2x_3 + 82.792x_1x_2x_3$$

$$\hat{\omega}_{\sigma} = 47.994 + 11.527x_1 + 15.323x_2 + 29.190x_3 + 29.566x_1x_2x_3.$$

$$\text{Min}\,(\hat{\omega}_{\mu} - 500)^2 + \hat{\omega}_{\sigma}^{2}$$

$$(x_1, x_2, x_3) = (1,\ 1,\ -0.525)$$

$$\hat{\omega}_{\mu} = 492.28$$

$$\hat{\omega}_{\sigma} = 44.01$$

Mean Squared Error (MSE) Criterion

$$\text{MSE} = \underbrace{(\hat{\omega}_{\mu}(\mathbf{x}) - \text{T})^2}_{\text{(bias)}} + \underbrace{(\hat{\omega}_{\sigma}(\mathbf{x}))^2}_{\text{(variance)}}$$

- (Lin and Tu, 1995)

$$\text{WMSE} = \lambda(\hat{\omega}_{\mu}(\mathbf{x}) - \text{T})^2 + (1-\lambda)(\hat{\omega}_{\sigma}(\mathbf{x}))^2$$

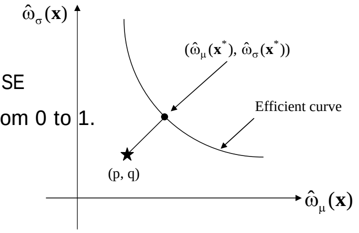
- T is the target value.
- λ is the weighting factor ($0 \le \lambda \le 1$).

➔ Lin and Tu (1995) implicitly set λ at 0.5.

Determination of λ in WMSE

Data-driven approach (Ding *et al.*, 2003)

- Procedure
  - Generate (p, q).
  - Find $\bar{\mathbf{x}}$'s that minimizes the WMSE criterion for various λ ranging from 0 to 1.
  - Plot the efficient curve for $\bar{\mathbf{x}}$'s.
  - Find $\mathbf{x}^*$ that minimizes $(\hat{\omega}_{\mu}(\bar{\mathbf{x}}) - \text{p})^2 + (\hat{\omega}_{\sigma}(\bar{\mathbf{x}}) - \text{q})^2.$
  - The corresponding λ^* is obtained.
- The meaning of λ^*
  - The weight at which the mean and standard deviation minimizing the WMSE criterion is closest to the ideal point.



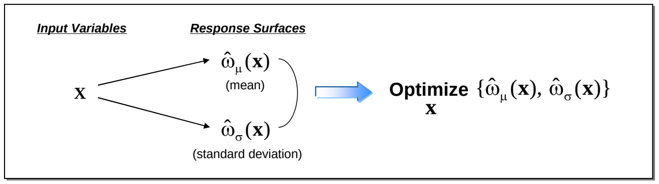
- (p, q): the ideal point of $(\hat{\omega}_{\mu}(\mathbf{x}^{\text{p}}), \hat{\omega}_{\sigma}(\mathbf{x}^{\text{q}}))$

$$\mathbf{x}^{\text{p}} = \arg\min_{\mathbf{x}} \{(\hat{\omega}_{\mu}(\mathbf{x}) - \text{T})^2\},\ \ \mathbf{x}^{\text{q}} = \arg\min_{\mathbf{x}} \{(\hat{\omega}_{\sigma}(\mathbf{x}))^2\}$$

Comparisons

Method	Optimal setting X*	$\hat{\omega}_{\mu}(X^*)$	$\hat{\omega}_{\sigma}(X^*)$	λ
VM	**(0.620, 0.230, 0.100)**	500.000	51.900	0.99*
LT	**(1.000, 0.074, -0.252)**	494.659	44.463	0.50
KL (d$_\mu$=-4.39, d$_\sigma$=0)	(1.000, 0 .086, -0.254)	496.111	44.632	0.58
CN (ρ=1, Δ=5)	**(0.975, 0.056, -0.214)**	495.020	44.727	0.52*
Proposed method	**(1.000, 0.089, -0.255)**	496.473	44.671	0.60

Dual Response Surface Problem



Input Variables

Response Surfaces

$\mathbf{x}$

$\hat{\omega}_{\mu}(\mathbf{x})$ (mean)

$\hat{\omega}_{\sigma}(\mathbf{x})$ (standard deviation)

Optimize $\{\hat{\omega}_{\mu}(\mathbf{x}), \hat{\omega}_{\sigma}(\mathbf{x})\}$ $\mathbf{x}$

Existing Works for Optimization Criterion

- **Vining and Myers (1990)**
  - Primary response
- **Lin and Tu (1995)**
  - Mean squared error (MSE)
- **Kim and Lin (1998)**
  - Minimum membership degree

Other Important Issues

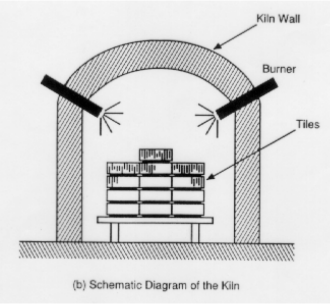
- *Combined Array vs. Product Array*
- *Type of Noise Factors*
- *Choice of Performance Measures*
- *The purposes of the inner array and the outer array are very different !!!*

Ina Tile Company (Japan, 1953)

Clay tiles fired in kiln

Problem: Size variation in tiles

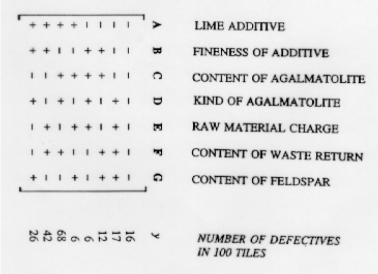
(Possibly



)

- Possible Remedies:
  - 1) Buy a new kiln with precise controllability of temperature and temperature gradient
  - 2) Seek a more “robust” recipe for tile clay

- Design of Experiment
  - Solution: Add 5% Lime
  - Size variation reduced
  - No additional cost

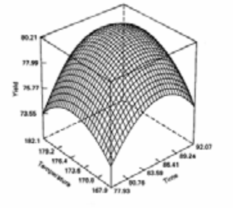


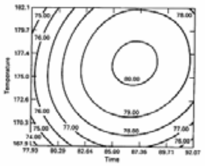
+	+	+	+	−	−	−	−	A	LIME ADDITIVE
+	+	−	−	+	+	−	−	B	FINENESS OF ADDITIVE
−	−	+	+	+	+	−	−	C	CONTENT OF AGALMATOLITE
+	−	+	−	+	−	+	−	D	KIND OF AGALMATOLITE
−	+	−	+	+	−	+	−	E	RAW MATERIAL CHARGE
−	+	+	−	−	+	+	−	F	CONTENT OF WASTE RETURN
+	−	−	+	−	+	+	−	G	CONTENT OF FELDSPAR
26	42	68	6	6	12	17	16	y	*NUMBER OF DEFECTIVES IN 100 TILES*

Multiple Response Optimization

RSM : Example

- y_1 = process yield (LTB)
- x_1 = reaction time, x_2 = reaction temperature
- Response Surface of y_1



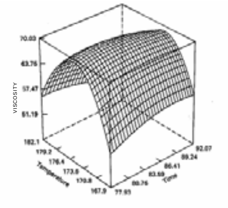


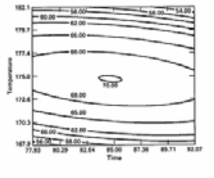
- $\hat{y}_1^* = 80.21$ at $\mathbf{x}^* = (x_1^*, x_2^*) = (87.0, 176.5)$

RSM : Example (continued)

- What if y_2(viscosity; NTB) and y_3(molecular weight; STB) are added?

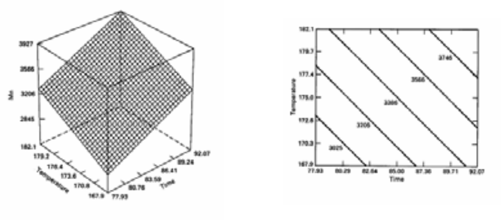
- Response Surface of y_2





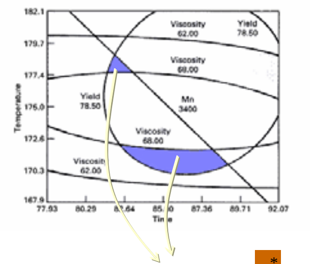
RSM : Example (continued)

- Response Surface of y_3



RSM : Example (continued)

- Acceptable Region : $y_1 \geq 78.5,\ 62 \leq y_2 \leq 68,\ y_3 \leq 3400$



Optimal $\mathbf{x}^* = (x_1^*, x_2^*) = ?$

Multi-Response System (MRS)

- Multiresponse variables measured for each setting of input variables
- Common problem in product/process design

Tire

y_1= PICO Abrasion Index
y_2= 200% Modulus
y_3= Elogation at Break
y_4= Hardness

x_1= Silica Level
x_2= Silane Coupling Agent level
x_3= Sulfur Level

(Derringer and Suich 1980)

Plasma-Etching Process

y_1= Linewidth of Track
y_2= Nonuniformity of Etch Rate
y_3= Etch Rate of Al/Si Alloy Layer
y_4= Photoresist Etch Rate
y_5= Oxide Etch Rate

x_1= Pressure of Reaction Chamber
x_2= Radio-frequency Power Level
x_3= Temperature
x_4= Flow Rate of Boron Trichloride
x_5= Flow Rate of Silicon Tetrachlorine
x_6= Flow Rate of Chlorine

(Logothetis and Haigh 1988)

Injection Molding of Washing Machine Agitator

y_1= Outer Diameter
y_2= Height
y_3= Pull-out Force

x_1= Mold Temperature
x_2= Injection Pressure
x_3= Hold on Pressure
x_4= Injection Time
x_5= Hold on Time
x_6= Cooling Time
x_7= Fill Time

(Reddy, Nishina, and Babu 1997)

Multi-Response System (continued)

- Dry Noodle Processing (Ventresca 1993)
- Beef Stew Pouch (Elsayed and Chen 1993)
- Hose-to-Connector Assembly (Goik, Liddy, and Taam 1994)
- Gear Hardening Process (Layne 1995)
- Bonding Process in Semiconductor (Del Castillo et al. 1996)
- Electrochemical Cutting Mechanism (Anjum et al. 1997)
- Characterization of Colloidal Gas Aphrons (Jauregi et al. 1997)

Examples of Multiple Response Surface Optimization Problems

- Quite often we may have an experimental optimization situation where we have two or more simultaneous responses for each factor configuration.

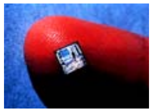
Examples:

➔ **Process:** HLPC assay optimization.
Responses: Resolution, Run time, Signal-to-noise, Tailing
Factors: pH, column temperature, etc.

➔ **Process:** Pharmaceutical tablet optimization .
Responses: dissolution rate, hardness, friability, etc
Factors: various excipient levels.

➔ **Process:** Chemical process optimization
Responses: Conversion, activity
Factors: time, temperature, and catalyst.

.



Examples of Multiple Response Surface Optimization Problems

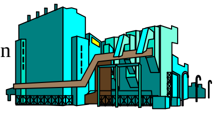
Process: Chemical Mechanical Polishing (semiconductor manufacturing)
Responses: Removal Rate, Non-uniformity (lack of flatness)
Factors: rotating speed, polish head down force, Back pressure.

Process: Etching process (semiconductor manufacturing)
Responses: etch thickness, etch uniformity (std. dev.)
Factors: Rotation speed, N_2 (nitrogen) flow, amount of oxide etched

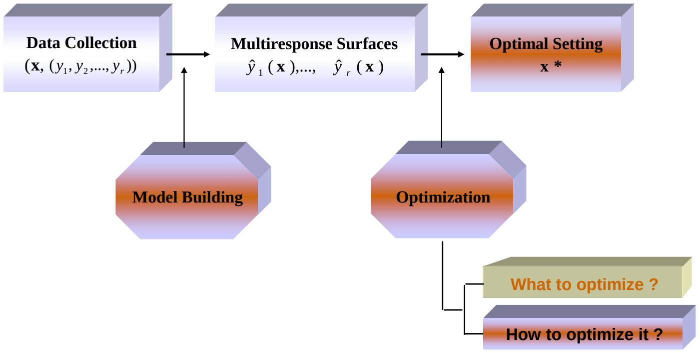
Process: Machining process (e.g. lathe used in metal cutting)
Responses: dimension(s) manufactured, surface finish, material removal ra
Factors: cut angle, feed rate, workpiece rotational speed

Process: Design of a force transducer
Responses: Nonlinearity and hysteresis
Factors: lozenge angle, bore diameter,
lozenge angle deviation, bore diameter deviation

Noise variables



MRS : Stages



Data Collection
$(\mathbf{x}, (y_1, y_2, \ldots, y_r))$

Multiresponse Surfaces
$\hat{y}_1(\mathbf{x}), \ldots, \hat{y}_r(\mathbf{x})$

Optimal Setting
$\mathbf{x}^*$

Model Building

Optimization

What to optimize ?

How to optimize it ?

Overlapping mean response surfaces

Responses:
Y_1 = particle size, Y_2 = glass transition temperature

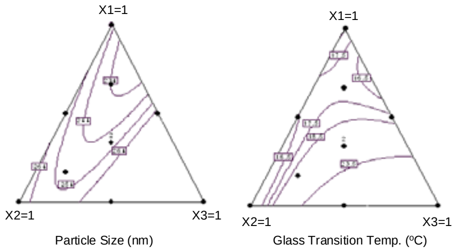
Factors:
x_1 = % of Pluronic F68, x_2 = % of polyoxyethylene sorbitan 40 monostearate,

and x_3 = % of polyoxyethylene sorbitan fatty acid ester NF
$(x_1 + x_2 + x_3 = 1 \text{ and } 0 \le x_i \le 1, \; i = 1,2,3)$

Goal: Choose factor levels such that they minimize Y_1 and Y_2 or at least keep $Y_1 \le 234$ and $Y_2 \le 18$.

Example: Pseudolatex formulation for a controlled drug release coating

Response Surfaces for Y_1 and Y_2 Means

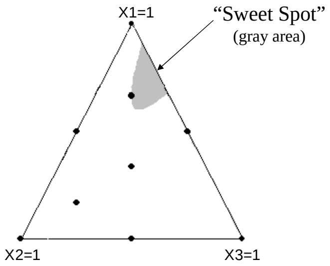


Design Expert Plots

Overlapping Contours Plots Produce a "Sweet Spot"

Here, the predicted response surface means are such that:

$\hat{E}(Y_1 \mid \mathbf{x}) \le 234$ and $\hat{E}(Y_2 \mid \mathbf{x}) \le 18$



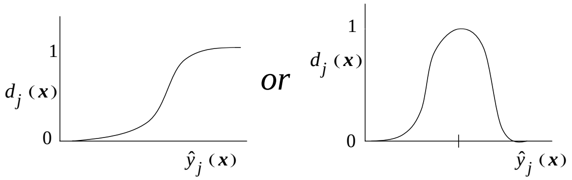
Design Expert Plot

Harrington (1965) or Derringer-Suich (1980) Desirability Functions

Currently multiple response surface optimization is done by using the following (geometric mean) objective function

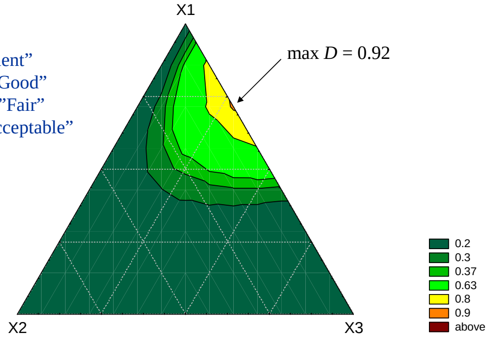
$$D(\mathbf{x}) = \left[\prod_{j=1}^{q} d_j(\mathbf{x}) \right]^{1/q},$$

where $d_j(\mathbf{x})$ is constructed to reflect the optimization desires of the experimenter for the j^{th} response type.



Contour Plot of the Harington Desirability Function for the controlled drug release coating example

Harrington Scale:
$D > 0.8$ => "Excellent"
$0.60 < D < 0.8$ =>"Good"
$0.37 < D < 0.60$ =>"Fair"
$D < 0.37$ => "Unacceptable"



max D = 0.92 for $\mathbf{x}$ = (0.75, 0, 0.25)

Drawbacks of the “sweet spot” and desirability function approaches.

- They do not take into account the underlying covariance structure of the multivariate response data.
- They do not take into account the uncertainty of the unknown model parameters.
- The “sweet spot” approach does not take into account how likely future values are to satisfy the experimenter’s specifications.
- The desirability approach does not take into account how likely future values are to satisfy the required desirability level.

How certain can we be that the “sweet spot” will produce sweet results?

In other words, what is $Pr(Y_1 \leq 234, Y_2 \leq 18 \mid \boldsymbol{x})$ for $\boldsymbol{x}$-points within the sweet spot?

By computing $p(\boldsymbol{x}) = Pr(Y_1 \leq 234, Y_2 \leq 18 \mid \boldsymbol{x})$ over the experimental region we can make such an assessment.

By taking a Bayesian approach it is possible to assess, in a straightforward way, the reliability of the conditional event

$$\{Y_1 \leq 234, Y_2 \leq 18\} \text{ given } \boldsymbol{X} = \boldsymbol{x}$$

in such a way as to take into account the uncertainty of the model parameters.

How often will the best Harrington desirability level be “at least good”?

- Since the desirability function is a mean it is mathematically possible in some cases that one or more of the responses could be out of spec. yet we still have an “acceptable” desirability level.
- Given the Harrington scale, one may want to consider

$$Pr(D(\boldsymbol{Y}) \geq 0.60 \mid \boldsymbol{x}) \quad \text{where} \quad \boldsymbol{Y} = (Y_1, Y_2)'$$

Harrington Scale:
$D > 0.8$ => “excellent”
$0.60 < D < 0.8$ =>”good”
$0.37 < D < 0.60$ =>”fair”
$D < 0.37$ => “very poor” to “poor”

Why is the “sweet spot” not so sweet?

- If the mean of Y at a point $\boldsymbol{x}$ is less than an upper bound, u, then all that guarantees is that $Pr(Y \leq u \mid \boldsymbol{x}) > 0.5$
- Suppose $\mu_{Y_1} \leq u_1$ and $\mu_{Y_2} \leq u_2$. If Y_1 and Y_2 were independent, then all that is guaranteed is that $Pr(Y_1 \leq u_1, Y_2 \leq u_2 \mid \boldsymbol{x}) > 0.25$

For k independent Y_i’s the situation becomes:

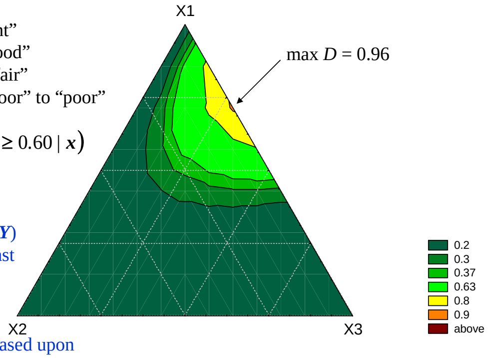
$$Pr(Y_1 \leq u_1, \ldots, Y_k \leq u_k \mid x) > 0.5^k$$

- If Y_1 and Y_2 are positively correlated then it may be easier to find $\boldsymbol{x}$-points to make $Pr(Y_1 \leq u_1, Y_2 \leq u_2 \mid \boldsymbol{x})$ large. Likewise, if Y_1 and Y_2 are negatively correlated (for each $\boldsymbol{x}$) then it may be more difficult.

Note: Corr(Y_1, $Y_2 \mid \boldsymbol{x}$) = -0.62 for the mixture experiment.

Contour Plot of the Harington Desirability Function for the controlled drug release coating example

Harrington Scale:
$D > 0.8$ => “excellent”
$0.60 < D < 0.8$ =>”good”
$0.37 < D < 0.60$ =>”fair”
$D < 0.37$ => “very poor” to “poor”



However, $Pr\left(D(\boldsymbol{Y}) \geq 0.60 \mid \boldsymbol{x}\right)$ is only 0.72 !

In other words, the chances of a new $D(\boldsymbol{Y})$ response being at least marginally “good” is only 72% despite the fact that the maximum D value based upon the **mean** response surfaces is rated “excellent”.

MRS Optimization : Approaches

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

* **dimensionality reduction strategy**

Priority – based Approach

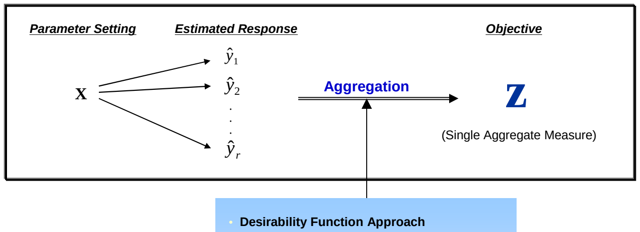
- Primary response vs. Secondary responses
- Framework

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

- 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)

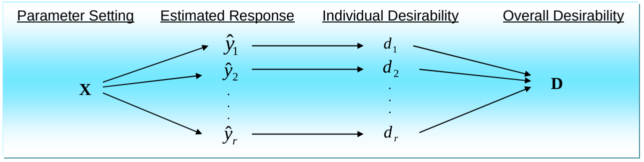
Framework of Dimensionality Reduction Strategy



- **Desirability Function Approach**
  Harrington(1965), Derringer and Suich(1980), Derringer(1994), Del Castillo, Montgomery, and McCarville(1996), Kim and Lin(2000)
- **Generalized Distance Approach**
  Church(1978), Khuri and Conlon(1981)
- **Loss Function Approach**
  Pignatiello(1993), Reibeiro and Elsayed(1994), Ames et al.(1997), Lin and Tu(1995), Vining(1998)

Desirability Function Approach

- Framework



Parameter Setting | Estimated Response | Individual Desirability | Overall Desirability

$\mathbf{X} \rightarrow \hat{y}_1, \hat{y}_2, \ldots, \hat{y}_r \rightarrow d_1, d_2, \ldots, d_r \rightarrow \mathbf{D}$

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

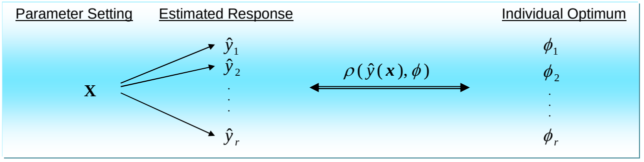
Harrington (1965)	Kim and Lin (1998)
Derringer and Suich (1980)	Kim and Lin (2000)
Derringer (1994)	Del Castillo, Montgomery, and McCarville (1996)
Goik, Liddy, and Taam (1994)	

Desirability Function Approach (cont’d)

- Derringer and Suich (1980), Derringer (1994)
  - $$d_j = \begin{cases} 0, & \hat{y}_j \le y_j^{\min}, \\ \left(\dfrac{\hat{y}_j - y_j^{\min}}{y_j^{\max} - y_j^{\min}} \right)^u, & y_j^{\min} < \hat{y}_j < y_j^{\max}, \\ 1, & \hat{y}_j \ge y_j^{\max}. \end{cases}$$
  - D = $(d_1^{\,w_1} d_2^{\,w_2} \ldots d_r^{\,w_r})^{1/\sum w_j}$
  - Hard to Interpret the Value of D

Generalized Distance Approach

- Framework



Parameter Setting | Estimated Response | Individual Optimum

$\mathbf{X} \rightarrow \hat{y}_1, \hat{y}_2, \ldots, \hat{y}_r \xleftrightarrow{\rho(\hat{y}(\mathbf{x}), \phi)} \phi_1, \phi_2, \ldots, \phi_r$

- $\rho(\hat{y}(\mathbf{x}), \phi)$= 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)

Generalized Distance Approach (cont’d)

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

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

where $\phi = [\phi_1, \phi_2, \ldots, \phi_r]'$ is the ideal optimum,

$\hat{\Sigma}$ is the estimator of the common variance-covariance matrix of the random errors $(\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_r)$,

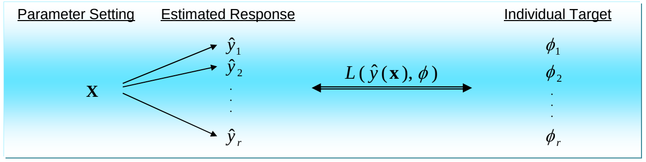
X is the design matrix, and

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.

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

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 “Mean” 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 :

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

$$\text{Such that} \quad -1 \le x_i \le 1 \qquad (i = 1,2,3)$$

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 “Standard Deviation” Models

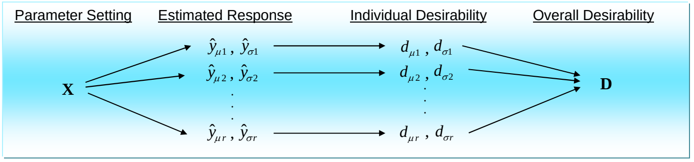
$\hat{y}_{\sigma 1}(\mathbf{x})$ = 0.06 + 0.11x_2 + 0.06x_3 + 0.12x_1^2 + 0.11x_3^2 - 0.10x_1x_3 + 0.05x_2x_3 (R^2 = 0.84)

$\hat{y}_{\sigma 2}(\mathbf{x})$ = 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



$\mathbf{D} = Minimum\ \{d_{\mu1},\cdots,d_{\mu r},\ d_{\sigma1},\cdots,d_{\sigma r}\}$ ➡ Find **X*** to Maximize **D**.

* Co-work with Kwang-Jae Kim

Proposed Approach : Formulation

$$\begin{aligned}
&\underset{x}{\text{Maximize}} \quad \lambda \\
&\text{subject to} \quad d_{\mu j}(\hat{y}_{\mu j}(\boldsymbol{x})) \geq \lambda, \quad j = 1, 2, \ldots, r, \\
&\qquad\qquad\quad d_{\sigma j}(\hat{y}_{\sigma j}(\boldsymbol{x})) \geq \lambda, \quad j = 1, 2, \ldots, r, \\
&\qquad\qquad\quad \mathbf{x} \in \Omega\ .
\end{aligned}$$

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
$T_{\mu j}$, $T_{\sigma j}$	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 $\boldsymbol{d} = (d_{\mu1}, d_{\mu2}, d_{\sigma1}, d_{\sigma2})$

    $\boldsymbol{d_1} = (0.5, 0.5, 0.5, 0.5)$ 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

Proposed Approach : Variations

V1 : Consideration of Alternative Responses

- Responses were alternatives rather than all being essential.

$$\underset{x}{\text{Maximize}}\{\ \text{maximum}\ (\lambda_1,\ \lambda_2,\ \ldots,\ \lambda_r)\}$$

subject to

$$d_{\mu_j}(\hat{y}_{\mu_j}(\boldsymbol{x})) \geq \lambda_j, \quad j = 1, 2, \ldots, r,$$

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

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

	Proposed Model	V1
x^*	$\begin{pmatrix}-0.21\\-0.40\\-1.00\end{pmatrix}$	$\begin{pmatrix}-0.03\\-0.04\\-1.00\end{pmatrix}$
$\hat{y}_{\mu 1}(d_{\mu 1})$	5.00 (0.50)	4.94 (0.48)
$\hat{y}_{\mu 2}(d_{\mu 2})$	0.37 (0.45)	0.38 (0.44)
$\hat{y}_{\mu 3}(d_{\mu 3})$	25.96 (0.73)	26.06 (0.74)
$\hat{y}_{\sigma 1}(d_{\sigma 1})$	0.06 (0.36)	0.10 (0.00)
$\hat{y}_{\sigma 2}(d_{\sigma 2})$	0.05 (0.50)	0.04 (0.57)
$\hat{y}_{\sigma 3}(d_{\sigma 3})$	1.64 (0.36)	1.26 (0.74)

Proposed Approach : Variations

V2 : Assignment of Different Weights on Mean and Standard Deviation

$$\underset{x}{\text{Maximize}} \quad \alpha\lambda_\mu + (1-\alpha)\lambda_\sigma$$

subject to

$$d_{\mu_j}(\hat{y}_{\mu_j}(x)) \geq \lambda_\mu\ , \quad j = 1, 2, \ldots, r,$$

$$d_{\sigma_j}(\hat{y}_{\sigma_j}(x)) \geq \lambda_\sigma\ , \quad j = 1, 2, \ldots, r,$$

$$\boldsymbol{x} \in \Omega\ ,$$

where $0 \leq \alpha \leq 1$.

	Proposed Model	V2		
		$\alpha = 0.1$	$\alpha = 0.5$	$\alpha = 0.9$
x^*	$\begin{pmatrix}-0.21\\-0.40\\-1.00\end{pmatrix}$	$\begin{pmatrix}-0.22\\-0.39\\-1.00\end{pmatrix}$	$\begin{pmatrix}-0.27\\-0.34\\-0.99\end{pmatrix}$	$\begin{pmatrix}-0.27\\-0.35\\-1.00\end{pmatrix}$
$\hat{y}_{\mu 1}(d_{\mu 1})$	5.00 (0.50)	4.98 (0.49)	4.91 (0.48)	4.91 (0.48)
$\hat{y}_{\mu 2}(d_{\mu 2})$	0.37 (0.45)	0.37 (0.46)	0.36 (0.48)	0.36 (0.48)
$\hat{y}_{\mu 3}(d_{\mu 3})$	25.96 (0.73)	25.98 (0.73)	26.08 (0.74)	26.06 (0.74)
$\hat{y}_{\sigma 1}(d_{\sigma 1})$	0.06 (0.36)	0.06 (0.36)	0.06 (0.35)	0.06 (0.36)
$\hat{y}_{\sigma 2}(d_{\sigma 2})$	0.05 (0.50)	0.05 (0.50)	0.05(0.50)	0.05 (0.50)
$\hat{y}_{\sigma 3}(d_{\sigma 3})$	1.64 (0.36)	1.64 (0.36)	1.65 (0.35)	1.64 (0.36)

Proposed Approach : Variations

V3 : Compensation of the “Maximin” Criterion

$$\underset{x}{\text{Maximize}} \quad [\lambda + \beta\sum_{j=1}^{r}(g_{\mu_j} + g_{\sigma_j})]$$

subject to

$$d_{\mu_j}(\hat{y}_{\mu_j}(\mathbf{x})) \geq \lambda, \qquad j = 1, 2, \ldots, r,$$

$$d_{\sigma_j}(\hat{y}_{\sigma_j}(\mathbf{x})) \geq \lambda, \qquad j = 1, 2, \ldots, r,$$

$$d_{\mu_j}(\hat{y}_{\mu_j}(\mathbf{x})) - g_{\mu_j} = \lambda, \quad j = 1, 2, \ldots, r,$$

$$d_{\sigma_j}(\hat{y}_{\sigma_j}(\mathbf{x})) - g_{\sigma_j} = \lambda, \quad j = 1, 2, \ldots, r,$$

$$\boldsymbol{x} \in \Omega\ ,$$

where $g_{\mu j}$ and $g_{\sigma j}$ are positive slacks;
β is a positive scaling constant.

	Proposed Model	V3		
		$\beta = 0.4$	$\beta = 0.7$	$\beta = 1.0$
x^*	$\begin{pmatrix}-0.21\\-0.40\\-1.00\end{pmatrix}$	$\begin{pmatrix}-0.21\\-0.40\\-1.00\end{pmatrix}$	$\begin{pmatrix}-0.20\\-0.36\\-1.00\end{pmatrix}$	$\begin{pmatrix}-0.09\\-0.23\\-1.00\end{pmatrix}$
$\hat{y}_{\mu 1}(d_{\mu 1})$	5.00 (0.50)	5.00 (0.50)	4.98 (0.50)	5.13 (0.53)
$\hat{y}_{\mu 2}(d_{\mu 2})$	0.37 (0.45)	0.38(0.45)	0.37 (0.45)	0.41 (0.39)
$\hat{y}_{\mu 3}(d_{\mu 3})$	25.96 (0.73)	25.95 (0.73)	25.99 (0.73)	25.88 (0.73)
$\hat{y}_{\sigma 1}(d_{\sigma 1})$	0.06 (0.36)	0.06 (0.36)	0.07 (0.33)	0.01 (0.00)
$\hat{y}_{\sigma 2}(d_{\sigma 2})$	0.05 (0.50)	0.05 (0.50)	0.05(0.51)	0.04 (0.56)
$\hat{y}_{\sigma 3}(d_{\sigma 3})$	1.64 (0.36)	1.64 (0.36)	1.57 (0.43)	1.15 (0.85)

MRS Optimization : Interactive Approach