Types of Experiments

Argriculture era (LNp.6)

LNp 11 **Treatment Comparisons :** Purpose is to compare several treatments of a factor (have 4 rice varieties and would like to see if they are different in terms of yield and drought resistence).

comparative experiment

Industrial era (LNp.6)

LNp 12 **Variable Screening :** Have a large number of factors, but only a few are important. Experiment should identify the important few.

→ factor sparsity principle

LNp 13 **Response Surface Exploration :** After important factors have been identified, their impact on the system is explored; regression model building.

$E(y_{\underset{\sim}{x}}) = f(x_1, \cdots, x_m) \approx x\beta$

Types of Experiments (Contd.)

If $E(y_{\underset{\sim}{x}}) = f(x_1, \cdots, x_m)$, what is $\arg \underset{\underset{\sim}{x}}{\max \atop \min} f(x_1, \cdots, x_m)$?

(f unknown)

LNp 14 **System Optimization :** Interested in determining the optimum conditions (e.g., maximize yield of semiconductor manufacturing or minimize defects).

Quality era (LNp.7)

LNp 15 **System Robustness :** Wish to optimize a system and also reduce the impact of uncontrollable (noise) factors. (e.g., would like cars to run well in different road conditions and different driving habits; an IC fabrication process to work well in different conditions of humidity and dust levels).

① ②

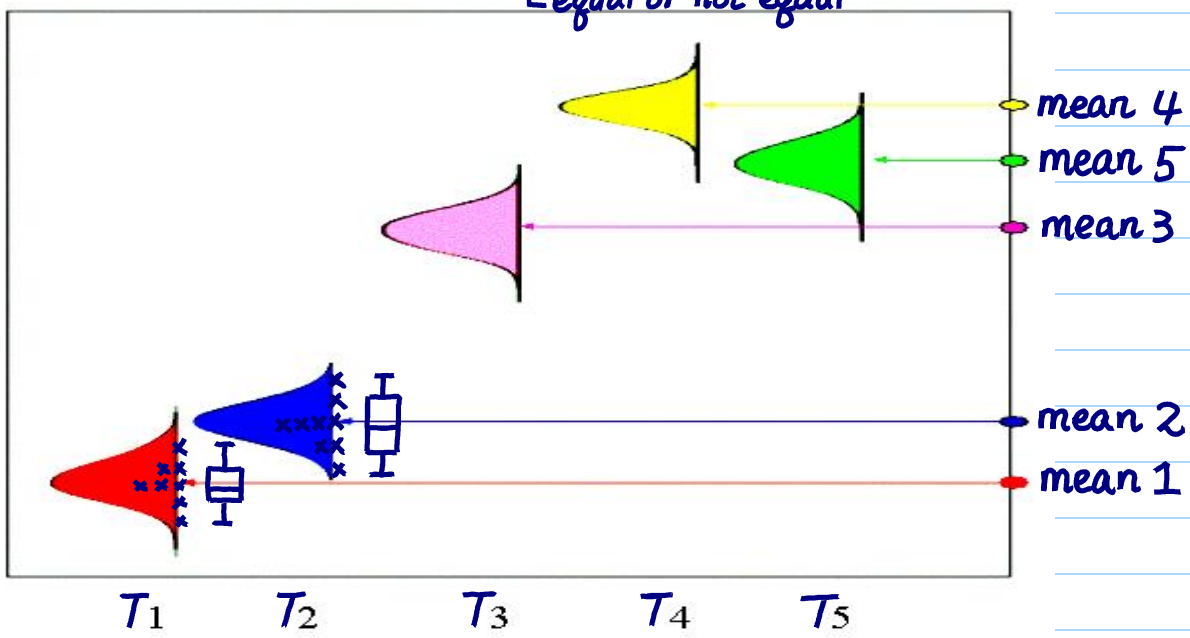
→ level (combinations) of noise factors

$E(y_{\underset{\sim}{x}}) = f(x_1, \cdots, x_m)$ ← for ①

$\mathrm{Var}_N(y_{\underset{\sim}{x}}) = h(x_1, \cdots, x_m)$ ← for ②

Treatment Comparison

Q: Whether the means are different? If yes, how different?
↑ equal or not equal



treatment: a combination of factor levels

Variable (Factor) Screening

effect

important effects

can use
- t- or F-test
- model selection

Objective: identify important factors or screen out unimportant factors

mean structure of $y_{\underset{\sim}{x}}$

$E(y_{\underset{\sim}{x}}) = f(x_1, \cdots, x_m) \approx f^*(x_1, x_3, x_8)$

factors

A, B, C, D, E → 2^5

effects

A_1, A_8

B_1, B_2, B_3

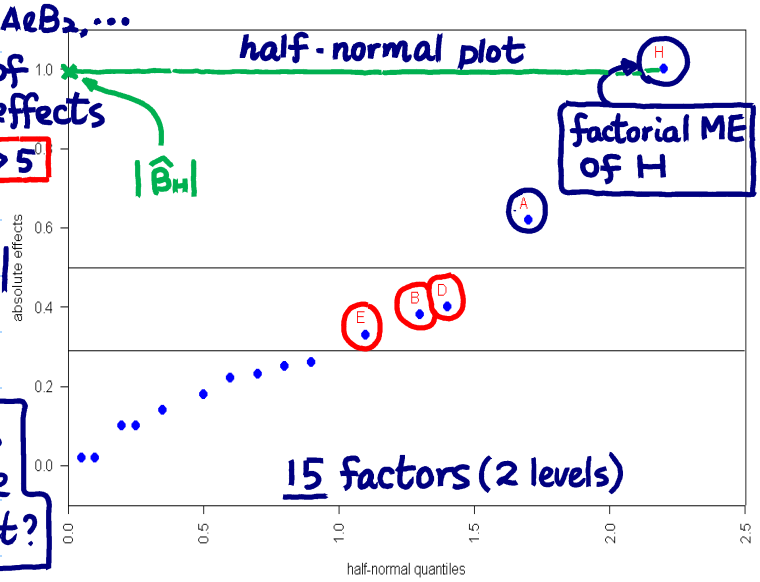
...

$A_8B_1, A_8B_2, \cdots$

u = # of all effects

$u \gg 5$

2^u



factorial ME of H

H, A: important

D, B, E: moderate

H+A: 50%

H+A+D+B+E: 65%

variation in y

Why is it more efficient?

When to use?: usually in the *preliminary* stage of the study of a system/process

main interest $\hat{y}_{\underset{\sim}{x}}$

Response Surface Approximation

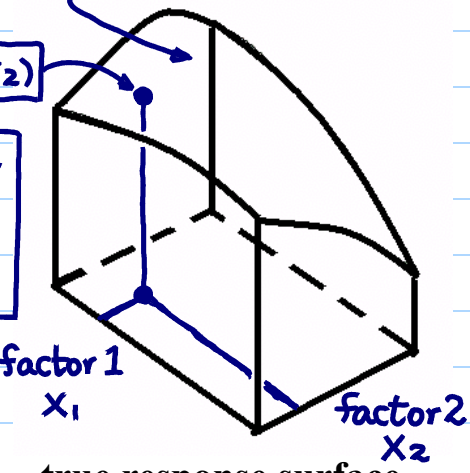
response surface: the relationship between a response and the factors $E(y_{\underset{\sim}{x}}) = f(x_1, \cdots, x_m)$

approximation $\hat{f}(x_1, \cdots, x_m)$

Objective:
develop a good approximation of the response surface

$E(y_{\underset{\sim}{x}})$

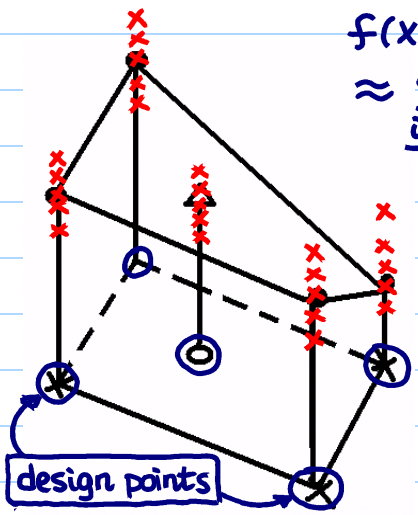
$f(x_1, x_2)$



true response surface

$f(x_1, \cdots, x_m) \approx \sum_{i=1}^{k} \hat{\beta}_i g_i(\underset{\sim}{x})$

use x (data) to estimate



approximate response surface

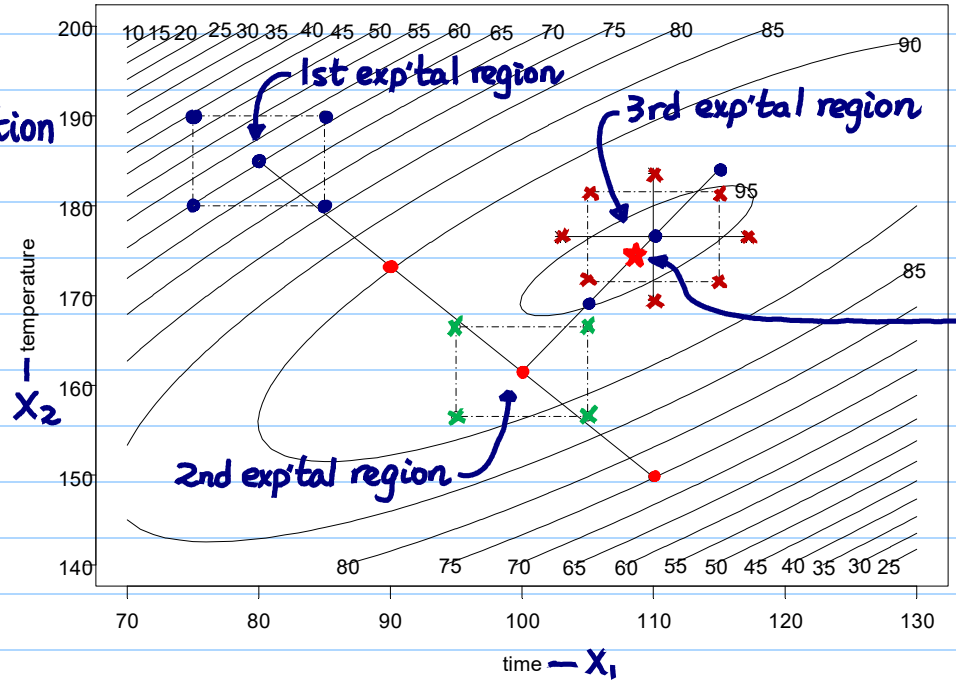
Recall 2 main objectives in regression:
① Predict $E(y_{\underset{\sim}{x}})$ or $y_{\underset{\sim}{x}}$
② Interpret $\underset{\sim}{\beta}$

System/Process Optimization

Executing Newton method using experiments

Objective: obtain optimal setting (of minimum/maximum response)

contour plot of true response surface $E(y_{\underset{\sim}{x}})$



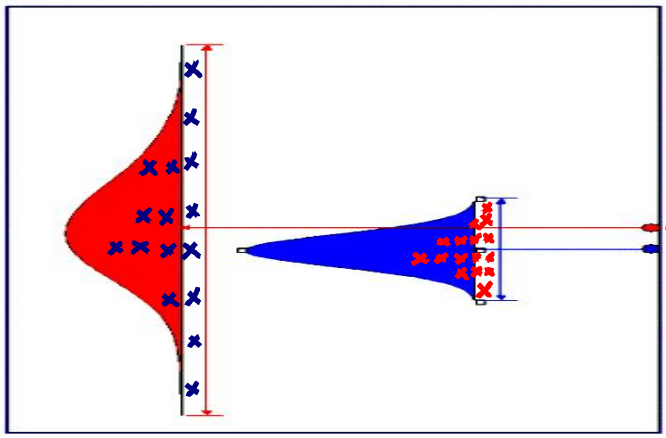
sequential experimentation

Q: where is the maximum of $E(y_{\underset{\sim}{x}})$?

achievable when $\underset{\sim}{X}$ & $\underset{\sim}{Z}$ have significant interaction(s)

Variation Reduction

Objective: adjust treatment factors ($\underset{\sim}{X}$) to make the system/process robust against noise variation ($\underset{\sim}{Z}$)



Q: which one will you choose?

Concept: *Besides optimizing the response, variation reduction is important in quality improvement.*

Systematic Approach to Experimentation

Problem formulation in statistical/mathematical language

e.g., treatment comparison	e.g., optimization
$y_{T_i} = \mu_i + \varepsilon$, $E(y_{T_i}) = \mu_i$	$y_{\underset{\sim}{x}} = f(x_1, \cdots, x_m) + \varepsilon$, $E(\varepsilon) = 0$
$H_0: \mu_1 = \mu_2 = \cdots = \mu_m$	$\arg\min_{\underset{\sim}{x}} f(x_1, \cdots, x_m)$ or $\arg\max_{\underset{\sim}{x}} f(x_1, \cdots, x_m)$

1. State the objective of the study.
2. Choose the response variable … should correspond to the purpose of the study.
   - Nominal-the-best, larger-the-better or smaller-the-better.

   (may use surrogate response; What information to be collected?; 望目 望大 望小)
3. Choose factors, levels, experimental region.
   - Use flow chart or cause-and-effect diagram.

   (identify possible "source of variation" in y; 2 levels ⇒ only linear effects; 3 levels ⇒ linear + quadratic effects)
4. Choose experimental design (i.e., plan).

   (how to choose design points in the space; a main focus of the course; design matrix → $x_1 \cdots x_m$: a level combination, a factor; cf)
5. Perform the experiment (use a planning matrix to determine the set of treatments and the order to be run).

   (e.g., randomization; $Y = X\beta + \varepsilon$: ⑤ Y, X ④; model matrix)
6. Analyze data (design should be selected to meet objective so that the analysis is efficient and easy).

   (e.g., orthogonality exists)
7. Draw conclusions.

   (offer solution to the problem in objective statement (e.g., optimization setting) → confirm exp't; follow-up exp'ts)