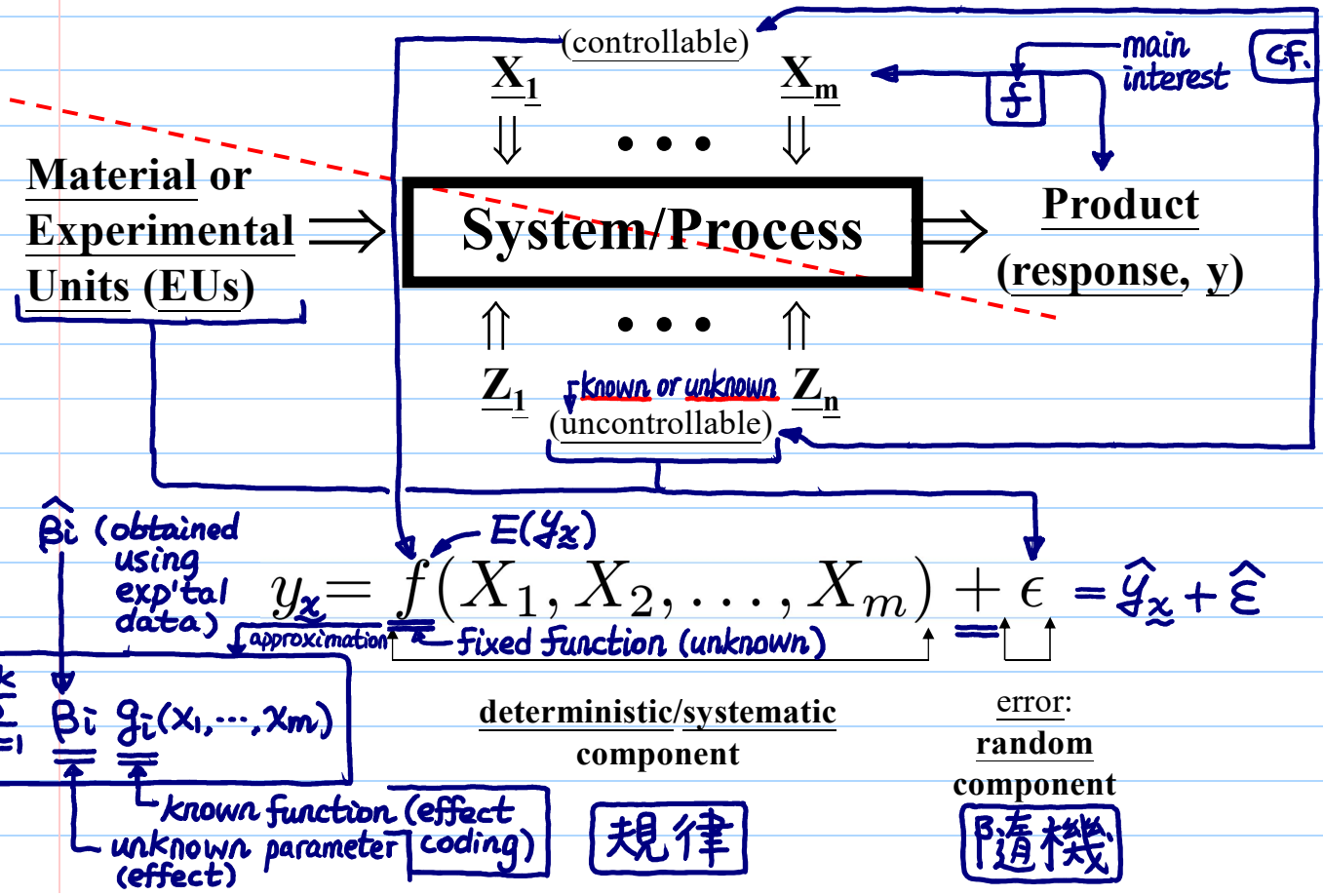
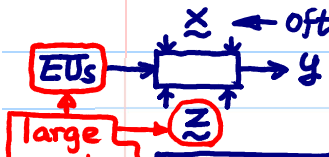
Conceptual model



Historical perspectives

- objectives in performing expts
- types of designs



生命

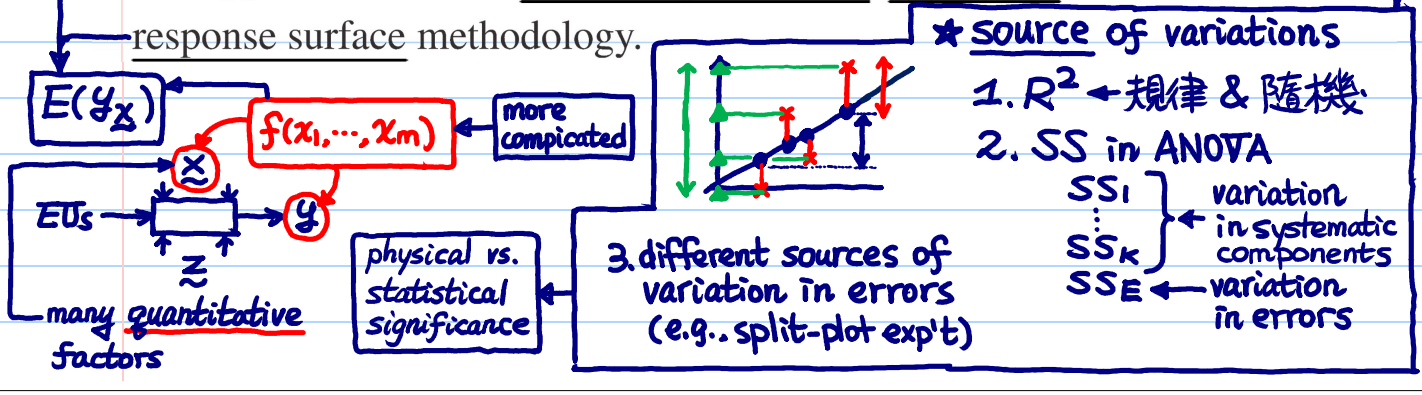
Agricultural Experiments : Comparisons and selection of varieties (and/or treatments) in the presence of uncontrollable field conditions, Fisher's pioneering work on design of experiments and analysis of variance (ANOVA).

e.g., replication, blocking, randomization & multiple comparison — examine whether there exist "difference" btwn treatments (comparative exp't)

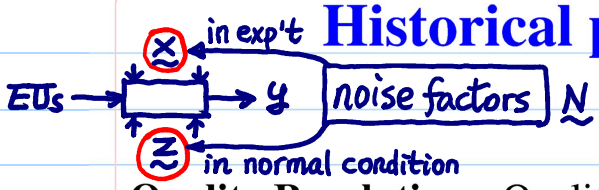
物

Industrial Era : Process modeling and optimization, Large batch of materials, large equipments, Box's work motivated in chemical industries and applicable to other processing industries, regression modeling and response surface methodology.

more x



Historical perspectives (Contd.)



- **Quality Revolution :** Quality and productivity improvement, variation reduction, total quality management, Taguchi's work on robust parameter design, Six-sigma movement.

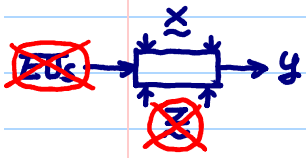
cf. study $E(y_x)$ in industrial era

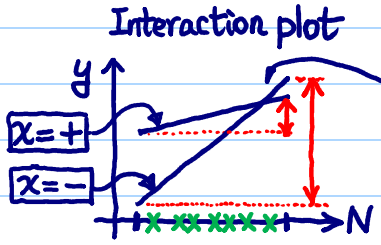
$E(y_x) = f(x_1, \cdots, x_m)$
$Var_N(y_x) = h(x_1, \cdots, x_m)$

- A lot of successful applications in manufacturing (cars, electronics, home appliances, etc.)

- **Current Trends and Potential New Areas :** Computer modelling and experiments, large and complex systems, applications to biotechnology, nanotechnology, material development, etc.

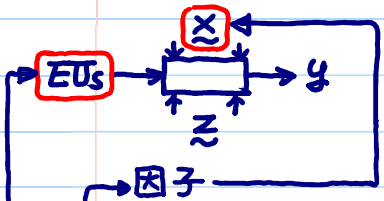
cf. physical experiment

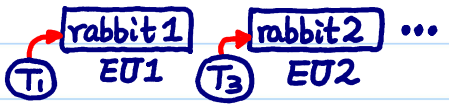


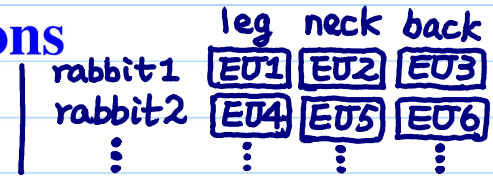


a strong interaction between X & N

Some Definitions



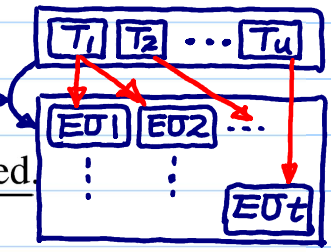




- **Factor :** variable whose influence upon a response variable is being studied in the experiment. temp.(溫度) : 20°C, 50°C ← 2 levels; pressure (压力) : 500, 1000, 1500 ppm ← 3 levels

- **Factor Level :** numerical values or settings for a factor. ← 水準, 水平

assignment



- **Experimental unit :** object to which a treatment is applied. ← 實驗單位

- **Trial** (or **run**) : application of a treatment to an experimental unit.

- **Treatment or level combination :** set of values for all factors in a trial.

T1 : (溫 20°C, 压 500 ppm) , T2 : (溫 50°C, 压 1500 ppm) , ...

- **Randomization :** using a chance mechanism to assign treatments to experimental units or run order.

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

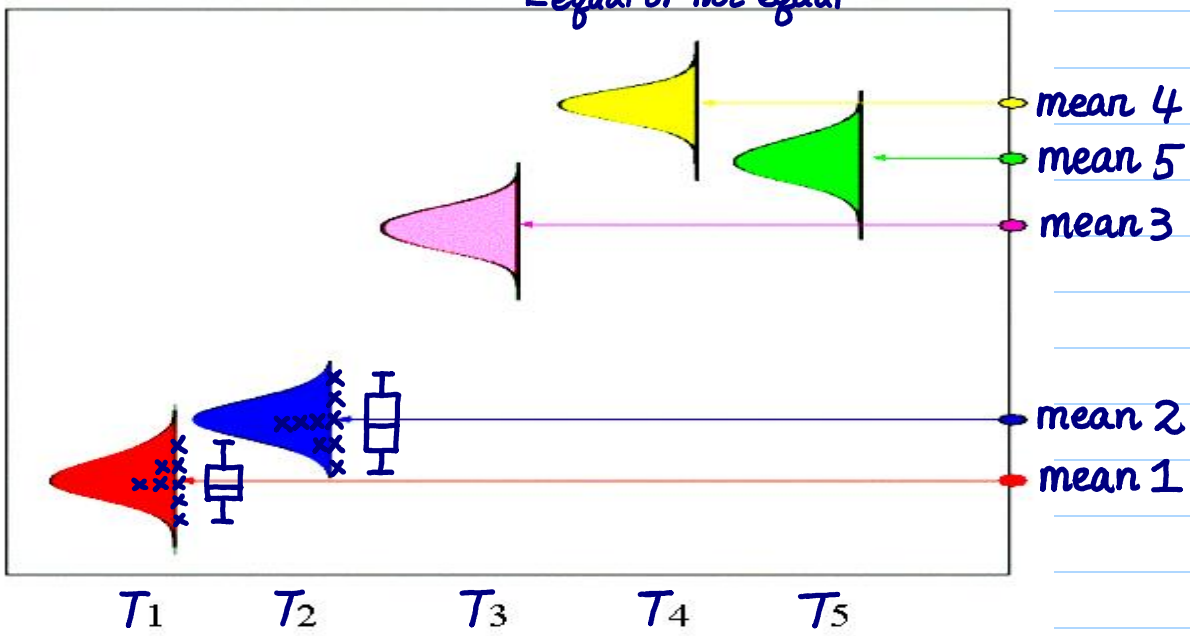
→ level (combinations) of noise factors

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

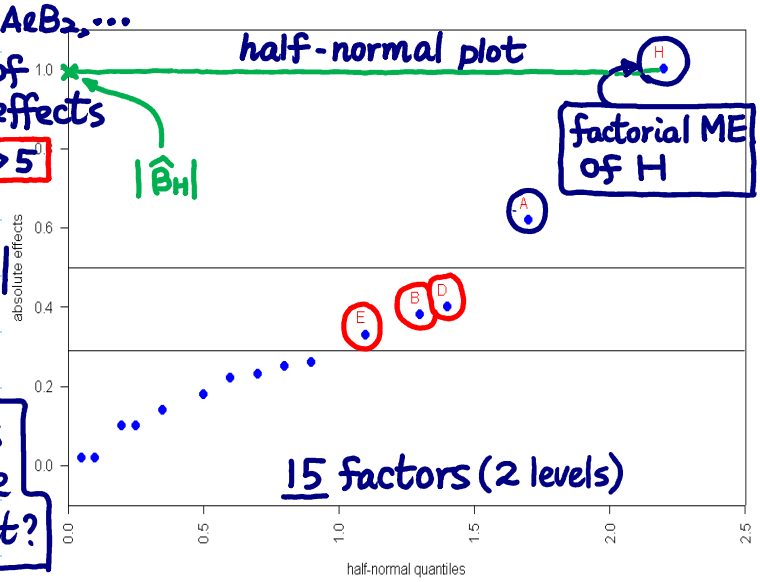
Objective: identify important factors or screen out unimportant factors

mean structure of $y_{\underline{x}}$
$E(y_{\underline{x}}) = f(x_1, \cdots, x_m) \approx f^*(x_1, x_3, x_8)$

factors
A, B, C, D, E → 2^5

effects
A, A_8
B_1, B_2, B_3
...
AeB_1, AeB_2, ...

u = # of all effects
$u \gg 5$
2^u



Why is it more efficient?

H, A: important

D, B, E: moderate

H+A: 50%

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

variation in y

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

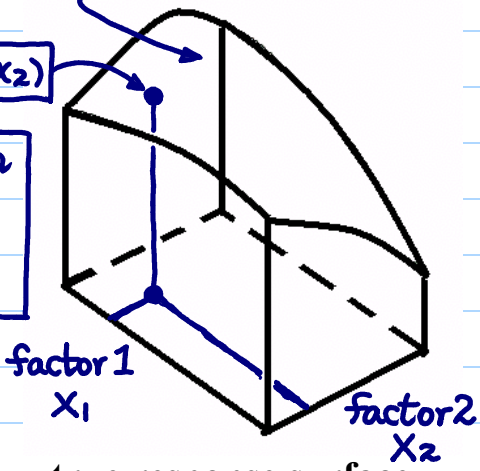
response surface: the relationship between a response and the factors $E(y_{\underset{\sim}{x}}) = f(x_1, \cdots, x_m)$

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

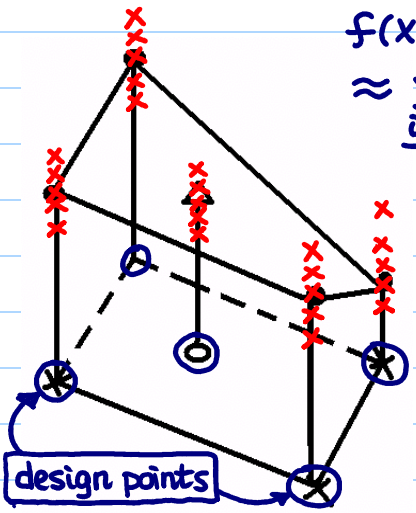
Objective:
develop a good approximation of the response surface

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

$f(x_1, x_2)$



true response surface



approximate 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

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

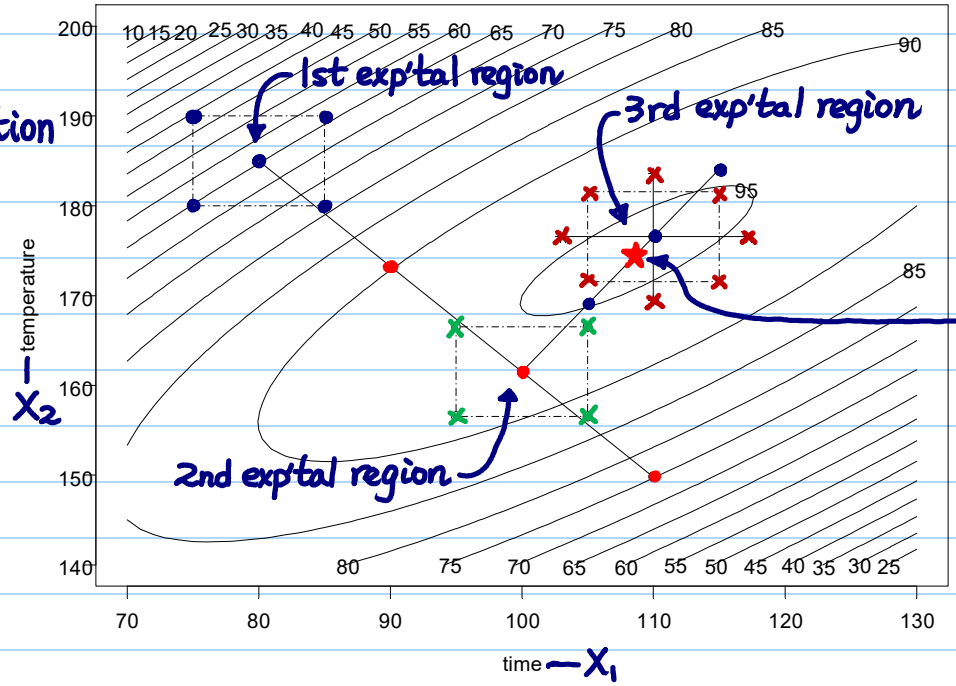
System/Process Optimization

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

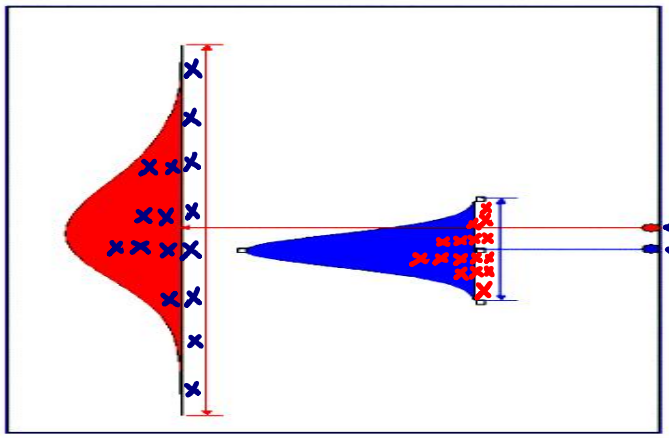


Q: where is the maximum of $E(y_{\underset{\sim}{x}})$?

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
$y_{Ti}=\mu_i+\varepsilon$, $E(y_{Ti})=\mu_i$	$y_{\underset{\sim}{X}}=f(x_1,\cdots,x_m)+\varepsilon$, $E(\varepsilon)=0$
$H_0: \mu_1=\mu_2=\cdots=\mu_m$	$\arg\min_{\underset{\sim}{X}} f(x_1,\cdots,x_m)$ 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 (望目), larger-the-better (望大) or smaller-the-better (望小).

   (may use surrogate response; What information to be collected?)
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 ↕ cf. $x_1 \cdots x_m$: a level combination, a factor)
5. Perform the experiment (use a planning matrix to determine the set of treatments and the order to be run). (e.g., randomization.)

   $Y = X\beta + \varepsilon$ (model matrix; ⑤ ∈ ④)
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)