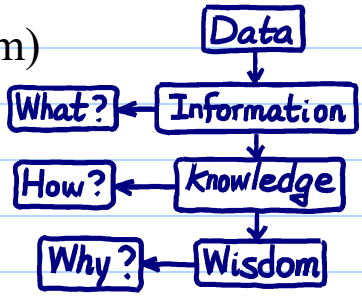
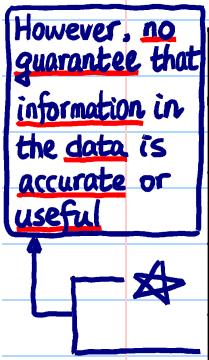
Question

What is Statistics?

- A branch of math --- calculation, derivative, proof, …
- A collection of many statistics (formula/algorithm)
- A useful tools for extracting information/knowledge from the data





哈利波特	Real Life
占卜學	Statistics
崔老妮	Statisticians
水晶球	Data
未來的資訊	Information

(卸貨)方法 → transformation

載體 → $\mathbb{R}^n$ (data) → $\mathbb{R}^1$, $\mathbb{R}^2$, ⋮, graphs

產品 → Information

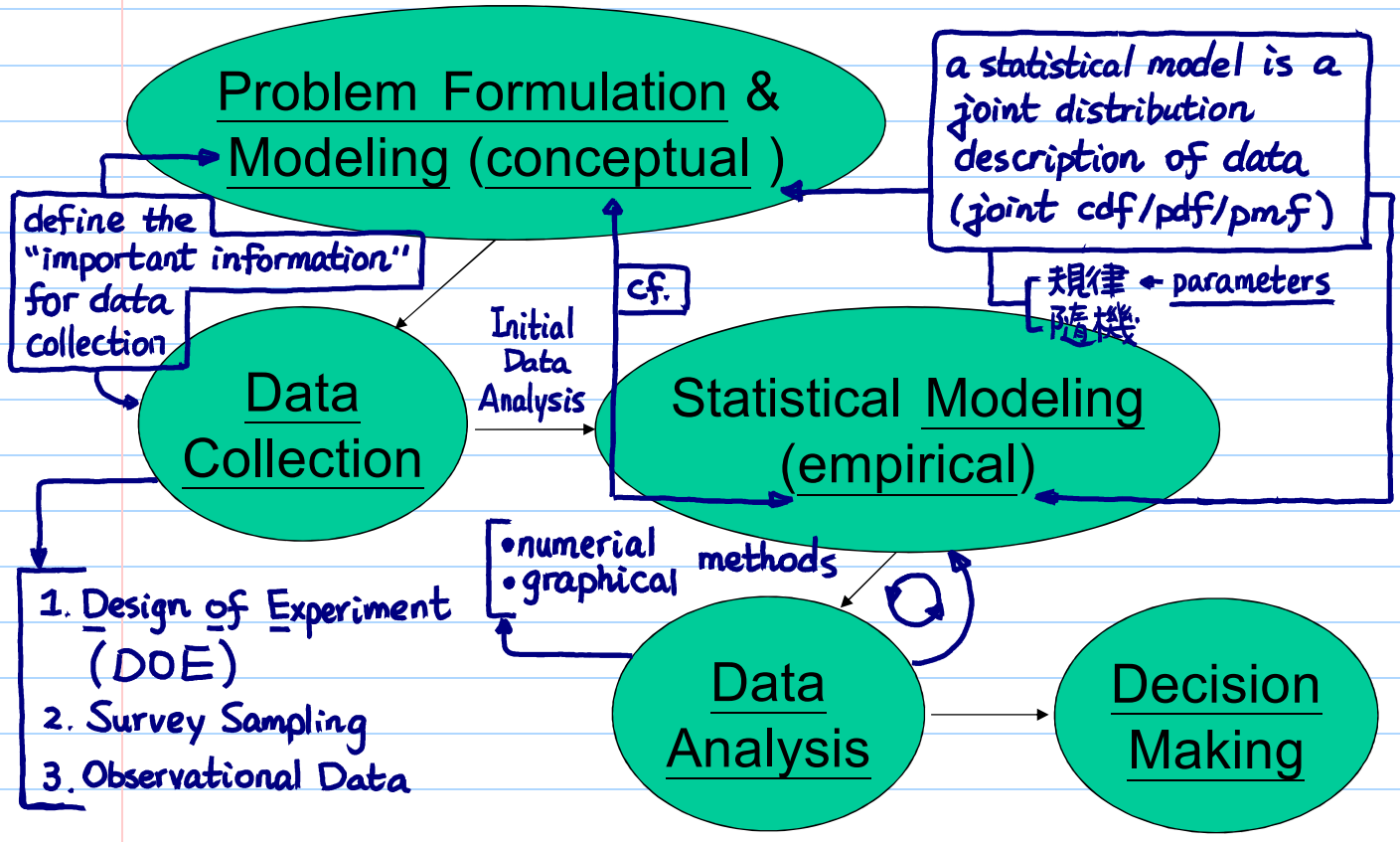
into ← ***aim of statistics***: provide ***insight*** by means of *data* → see, watch

对现象、状况、系统、...

Basic Procedures of Statistics

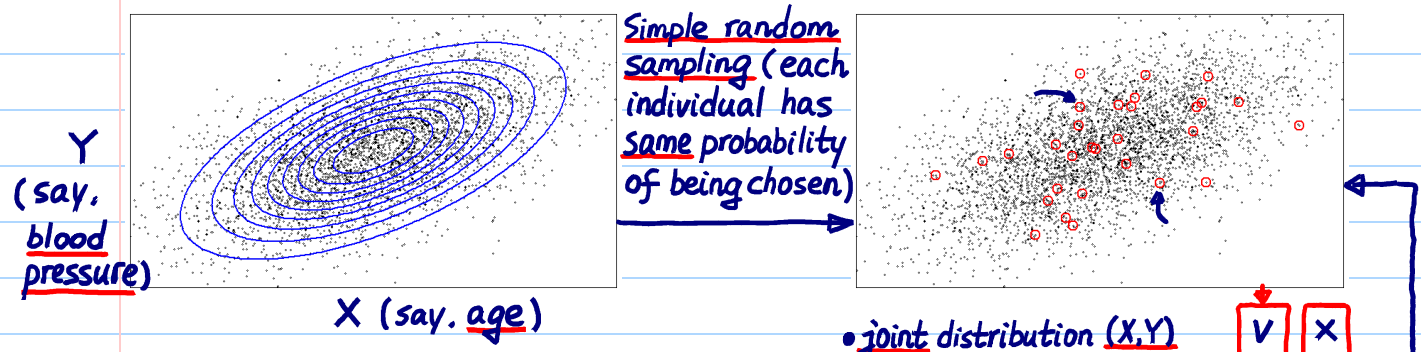
Recall Linear Models (STAT 5410) LN p1-2 ~ 12

- Statistics divides the study of data into *five* steps:



Distinction between Survey Sampling and DOE

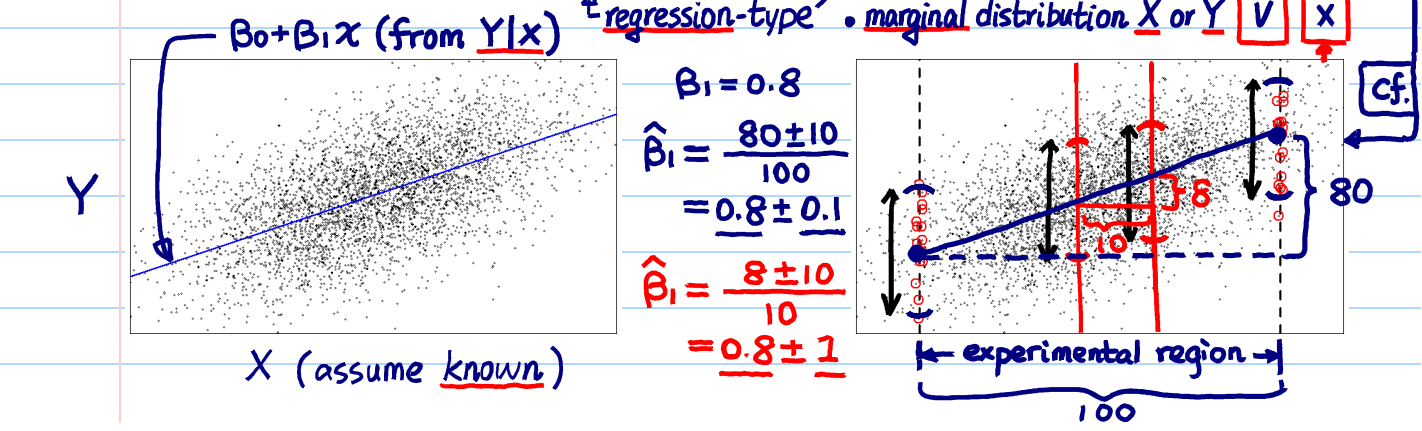
- Survey sampling: interested in "distribution"



- DOE: interested in "relationship"

regression-type

- joint distribution (X,Y) V X
- conditional distribution Y|x V VV
- marginal distribution X or Y V X



how to collect data such that 規律 is more apparent when 隨機 exist

Experimentation

- Experimentation: one of the most common activities that people engage in
  - ➢<神農本草經>: trial-and-error experiment
    "神農嘗百草，日遇七十二毒，得荼而解之"
  - ➢Experiment: a learning process for
    - knowledge gathering
    - problem solving
    - conjecture testing
- Modern experimental design dates back to 1930s (systematic approach)
- The results of an experiment often contains two parts:
  - 規律 ➢ *deterministic* component
  - 隨機 ➢ *random* component

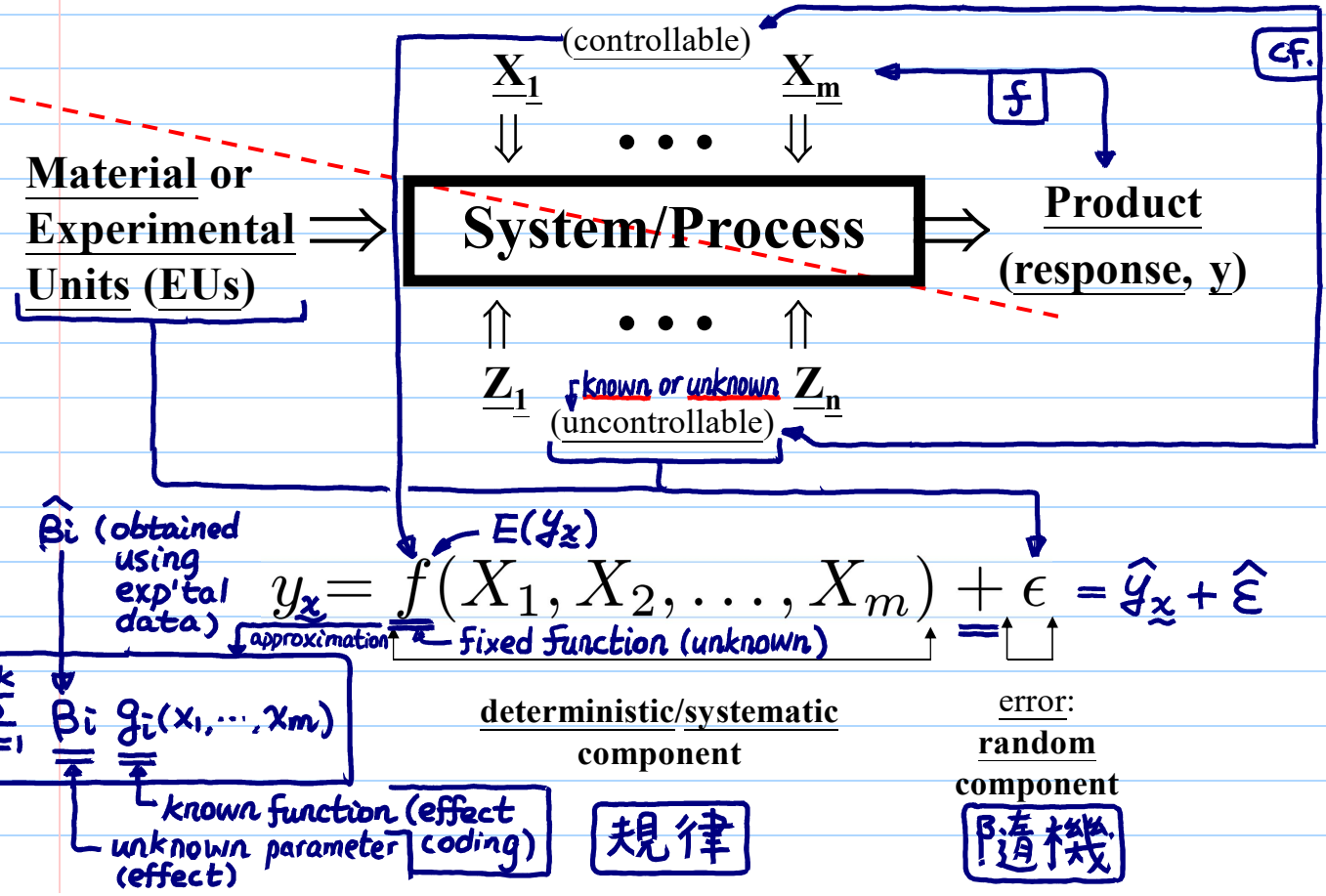
- DOE
- DAE (Design & Analysis of Experiments)

both components appear in statistical models (conceptual & empirical), e.g.,

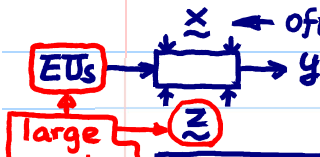
$$y_x = f(x_1, \cdots, x_m) + \varepsilon = \hat{y}_x + \hat{\varepsilon}$$

overfitting or lack of fit ?

Conceptual model



Historical perspectives



* objectives in performing exp'ts
* types of designs

causing large variation in EUs & z → large Var(ε)

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).

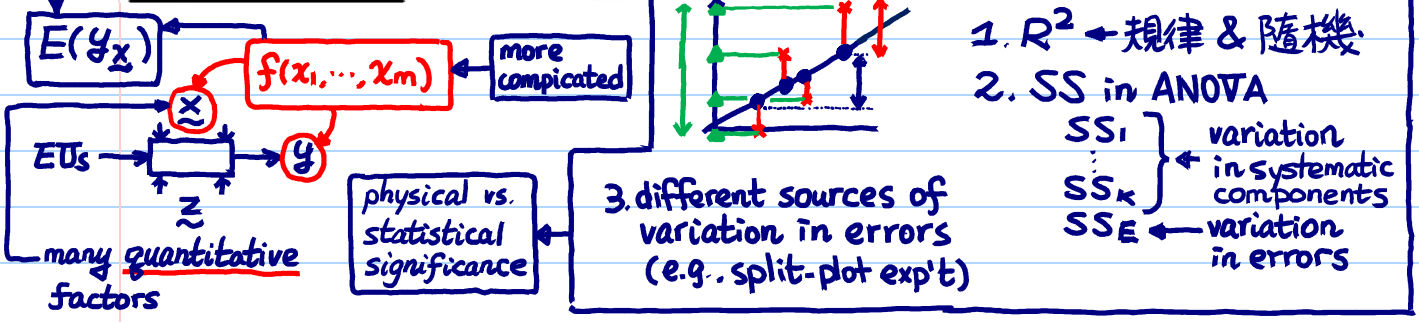
e.g., replication, blocking, randomization & multiple comparison

examine whether there exist "difference" btwn treatments (comparative exp't)

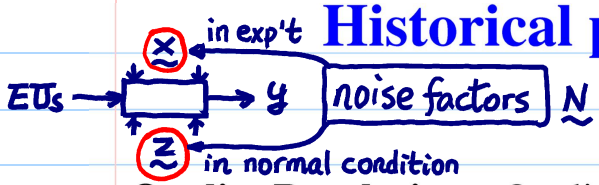
物

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.

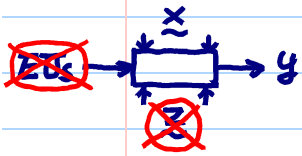
cf. study $E(y_{xz})$ in industrial era

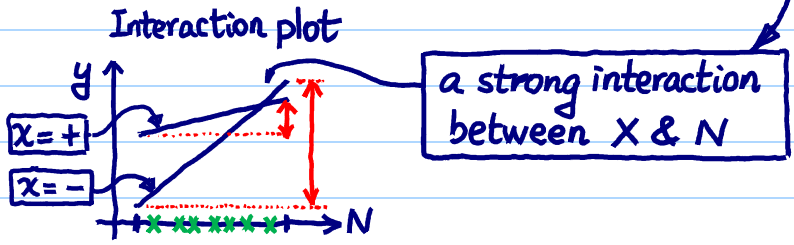
$E(y_{xz}) = f(x_1, \cdots, x_m)$
$Var_N(y_{xz}) = h(x_1, \cdots, x_m)$

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

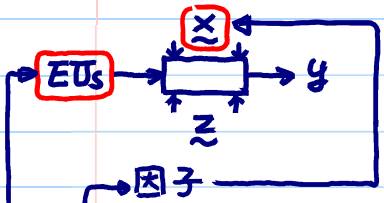
- **Current Trends and Potential New Areas :** Computer modelling and experiments, large and complex systems, applications to biotechnology, nanotechnology, material development, etc.

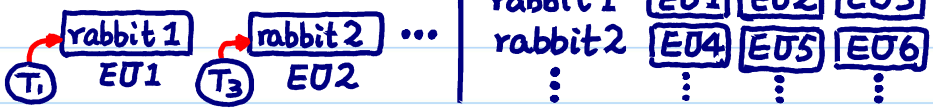
cf physical experiment





Some Definitions

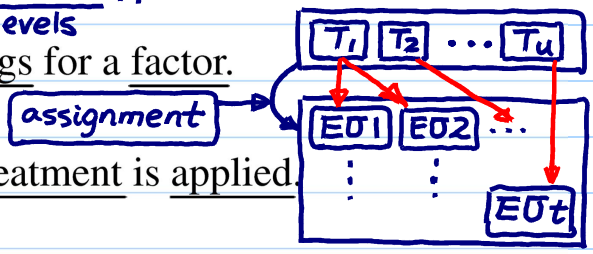




- **Factor :** variable whose influence upon a response variable is being studied in the experiment. temp.(溫度) : 20°C, 50°C ← 2 levels; pressure (压力) : 500, 1000, 1500 ppm ← 3 levels

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

- **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.

T_1 : (溫 20°C, 压 500 ppm), T_2 : (溫 50°C, 压 1500 ppm), ...

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