공정 모형 및 해석
실험계획법

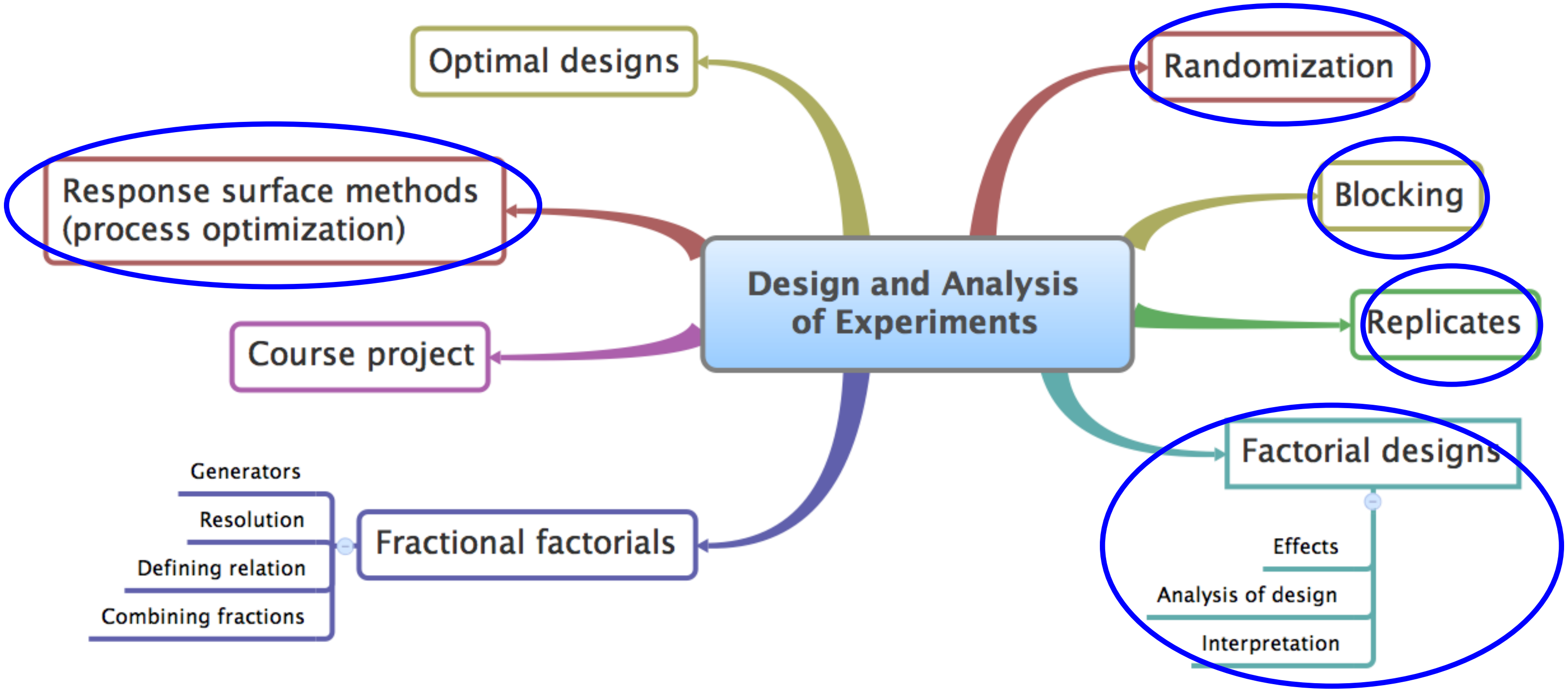
Jay Liu

Dept. Chemical Engineering

PKNU

Design of Experiments

- What we will cover



Reading:
http://www.chemometrics.se/index.php?option=com_content&task=view&id=18&Itemid=27

Usage examples

- Colleague: 8 factors seem to affect melt index. How to narrow them down? Which one has most effect on y?
- Engineer: 3 factors of interest; how to run the experiments?
- Manager: how do we analyze experimental data to optimize our process?
- Colleague: small changes in the flowrate lead to unsafe operation. Where can we operate to get similar results, but more safely?

Why design?

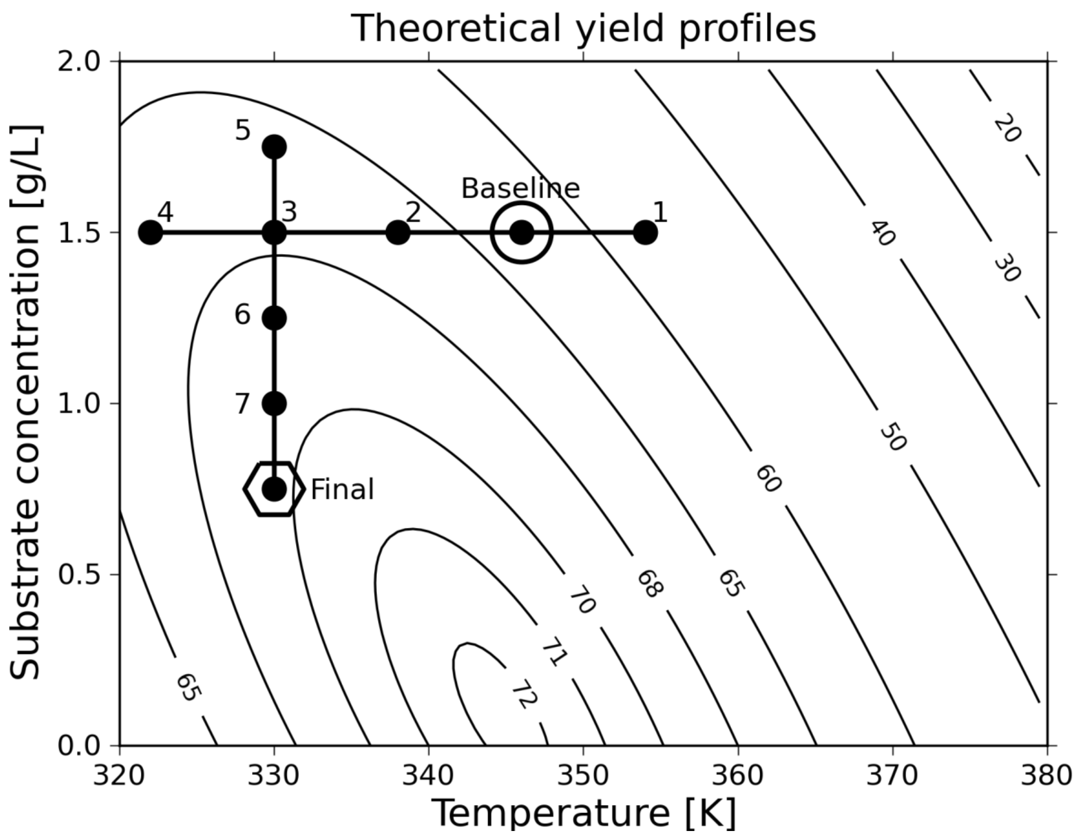
1. **Ensure adequate variability in all key variables.**
   - Variable *x* may have very important effect on process performance.
   - But if variation in it is small relative to noise level, then may
     - Accept H_0: effect of $x = 0$
     - Obtain confidence interval on effect of *x* to include zero.
   - This does not necessarily mean that effect of *x* is not important – only that it isn't large enough in this particular data set to detect significance.
   - Design of experiments provides a form of guarantee that accepting H_0 implies that the effect is not important.

Why design?

2. **Ensure identifiability of all important effects & interactions**
   - DOE helps ensure that all important main effects and interaction can be identified – minimizes confounding
   - Our bad experimental habits arise from the nature of university laboratories:
     - These undergrad labs aimed at demonstrating theoretical principles, not a building models, exploring for unknown effects, or optimizing processes.
     - Ex. Demonstrate the effect of temp. on reaction equilibrium – changing temp. holding all other variables constant!
   - COST approach is not good when searching for effects, building models, or optimizing processes.

[FYI]Changing One variable at a Single Time (COST)

- We can hardly find values of conc. & temp. for max. yield using COST approach



  - DOE: efficient ways of changing many variables at once

Why design?

3. **Maximize the information obtained in fewest number of experiments**
   - Examples of industrial screening experiment
     - Problem: in a new plant the cycle time in the filtration section was unacceptably long.
     - Need to de-bottleneck
     - Many factors suggested that might be responsible.
     - How to screen out important ones in fewest runs possible?
4. **Distinguish between causality and correlation**
   - Data from Australia over many years on
     - # of Baptist minister
       vs. amount of liquor consumed
     - Strong correlation? Causal effect?

Analysis of effects of a single variable at two levels

- Simplest case:
  - catalyst A vs catalyst B
  - low RPM vs high RPM
  - Etc
- Measure n_A value from setup A
- Measure n_B values from setup B
- Hold all other variables constant (control disturbances)

➔ Two ways to answer this:

- Comparing means of X and Y
- Least squares

Using confidence interval of $\bar{X} - \bar{Y}$

- Test for difference

$$s_p^2 = \frac{(n_A - 1)s_A^2 + (n_B - 1)s_B^2}{n_A - 1 + n_B - 1} \qquad \frac{(\bar{X} - \bar{Y})}{\sqrt{\frac{s_p^2}{n_A} + \frac{s_p^2}{n_B}}} \sim t_{n_A + n_B - 2}$$

- Confidence interval

$$\left[(\bar{X} - \bar{Y}) - t_{n_A + n_B - 2, \alpha/2} \sqrt{\frac{s_p^2}{n_A} + \frac{s_p^2}{n_B}}, (\bar{X} - \bar{Y}) + t_{n_A + n_B - 2, \alpha/2} \sqrt{\frac{s_p^2}{n_A} + \frac{s_p^2}{n_B}} \right]$$

Using least squares

- The same result can be achieved using least squares: $y_i = a_o + a_1 d_i$
  - d_i = 0 for A; d_i = 1 for B; y_i : the response variable

EXAMPLE (No. 6 in mid-term exam): Etch rate of solutions 1 & 2

- C. I approach

$$(\bar{x}_1 - \bar{x}_2) - t_{\alpha/2, n_1+n_2-2}(s_p)\sqrt{\frac{1}{n_1}+\frac{1}{n_2}} \le \mu_1 - \mu_2 \le (\bar{x}_1 - \bar{x}_2) + t_{\alpha/2, n_1+n_2-2}(s_p)\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}$$

$$(9.97 - 10.4) - 2.101(.340)\sqrt{\frac{1}{10}+\frac{1}{10}} \le \mu_1 - \mu_2 \le (9.97 - 10.4) + 2.101(.340)\sqrt{\frac{1}{10}+\frac{1}{10}}$$

$$-0.749 \le \mu_1 - \mu_2 \le -0.111$$

Zero included?

- LS approach

	value	S.E	t statistic	P-value	L.B 95%	U.B 95%
a0	9.97	0.107523	92.72469	1.41E-25	9.744103	10.1959
a1	0.43	0.15206	2.827832	0.011151	0.110534	0.749466

Confidence intervals of a_0 & a_1

Zero included?

- Same result and more (**significance test + prediction model**)

Concepts in DOE

Randomization and blocking

- Comparative experiment: effect of two methods on strength of rubber strip
  - Run experiments   Run order →

A	A	A	…	A	B	B	B	…	B

  and do significance test (C.I of $\bar{X}_A - \bar{X}_B$) or least squares($y_i = a_o + a_1 d_i$)

  ….. Any problem with this?

  - What if strip of rubber had variation along its length?

  Then, $\bar{X}_A - \bar{X}_B$ might just be reflecting this difference.

  - One solution → randomize allocation of rubber portion to methods (A&B)

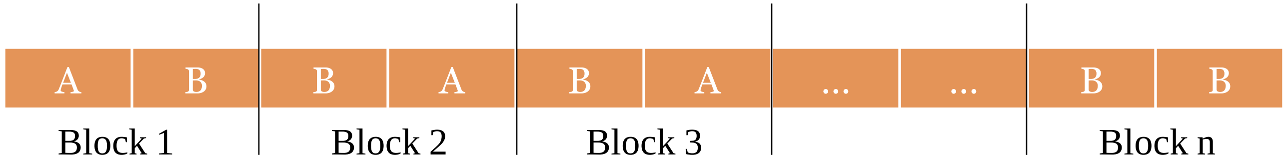
A	B	B	A	B	A	…	A	B	B

  …… No problem with this?

Concepts in DOE - Randomization and blocking

- Suppose we expect variation in rubber to be progressive along length of the strip! Then, two different adjacent portion will be much more similar than two distant ones.

  → **block into pairs** of adjacent pieces. Assign methods (A&B) randomly within block .



(Randomized block design)

And only compare within block

block	A B	$D = X_A - X_B$
1	X_{A1} X_{B1}	$d_1 = X_{A1} - X_{B1}$
2	X_{A2} X_{B2}	d_2
…	…	…
n	X_{An} X_{Bn}	d_n

Blocking can remove effect of possible uncontrolled variations along the length of strip (**remember advantage of paring**)