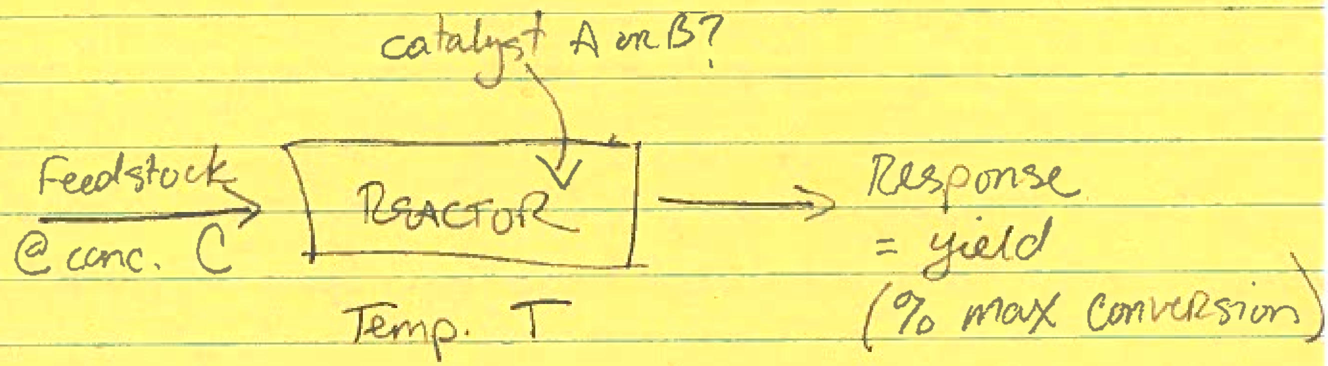
Spinning up A PILOT PLANT



Goal: find best operating conditions.

We consider all 3 factors.
What else are we ignoring?

- Pressure
- flow rate (residence time)

Asking only if catalyst A better than B requires

- fixing all other variables
- ignoring interactions b/w variables.

Factorial Designs

OFAT = one factor at a time

Why?

1. OFAT assumes all other factors don't matter
2. OFAT does not generalize (ie. not robust)

Factor = independent variable

multi-factorial = >1 factor

Response = output under study

= dependent variable

We won't cover multi-response

During experiment, factors are set at levels

Focus on 2-level to start

- Smaller experiments
- quant & qual variables handled the same

2-level factors A, B, C, ... have levels

$+1 \equiv + \equiv$ "high"

$-1 \equiv - \equiv$ "low"

For continuous/quant. factors, − level is usually the smaller number.

For qual factors, levels are arbitrary

- If we have a "baseline" usually we set this −. (wt, no drug)

Treatment = unique combination of factor levels
≈ "condition"

Run = operation of the system set at a particular treatment
≈ "sample"

Replicates = repeated runs at same treatment
≈ "biological replicate"

Replicates should not be run sequentially
Ideally, apparatus/system reset b/w runs

duplicates = repeated measurements of the same run
≈ "technical replicates"

- duplicates should be averaged <u>before</u> data analysis
- do not count toward degrees of freedom

1 run per treatment = "<u>un</u>replicated"
>1 run " " = "replicated"

OFAT design

		−	+
Temprature	T(°C)	160	180
Concentration	C(%)	20	40
Catalyst	K	A	B

coded units

design matrix

T	C	K	yield (y)
−	−	−	60
+	−	−	72
−	+	−	54
−	−	+	52

contrast

planning matrix

T(°C)	C(%)	K(A or B)	yield (%)
160	20	A	60
180	20	A	72
160	40	A	54
160	20	B	52

Replicating the −,−,− would add a run but not a treatment.

Problem w/ OFAT:
- Everything compared to baseline, but what if there is no baseline?
  - New process
  - Genetic variants

4

EXAMPLE: Box, Hunter, & Hunter, 2005 (BHH)

Table 5.2. A 2^3 Factorial Design: Pilot Plant Investigation

Temperature, T (°C)		Concentration, C (%)		Catalyst, K		3 factors
−	+	−	+	−	+	"coded units"
160	180	20	40	*A*	*B*	2 levels

	Coded Units of Factors			
	T	**C**	**K**	Average Yield *y* from Duplicate Runs
	−	−	−	60
	+	−	−	72
	−	+	−	54
	+	+	−	68
	−	−	+	52
	+	−	+	83
	−	+	+	45
	+	+	+	80
Run Number	Temperature, **T** (°C)	Concentration, **C** (%)	Catalyst, **K** (*A* or *B*)	Yield, **y** (%)
	Operational Levels of Factors			
1	160	20	*A*	60
2	180	20	*A*	72
3	160	40	*A*	54
4	180	40	*A*	68
5	160	20	*B*	52
6	180	20	*B*	83
7	160	40	*B*	45
8	180	40	*B*	80

[a] The data were from a real example that has, however, been considerably simplified to allow us to concentrate on important issues and avoid distracting fractions and decimals.

column is called a "contrast"

design matrix (excluding y)

Runs done in random order IRL.

planning matrix (excluding y)

5

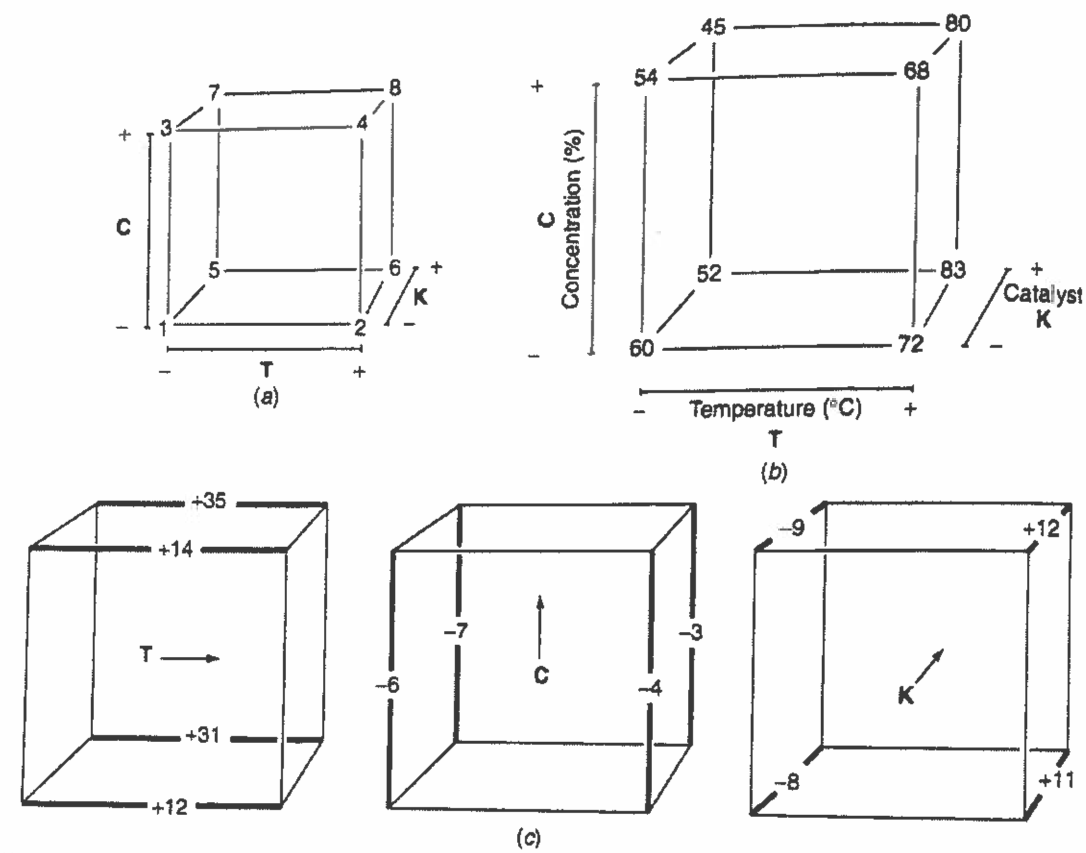
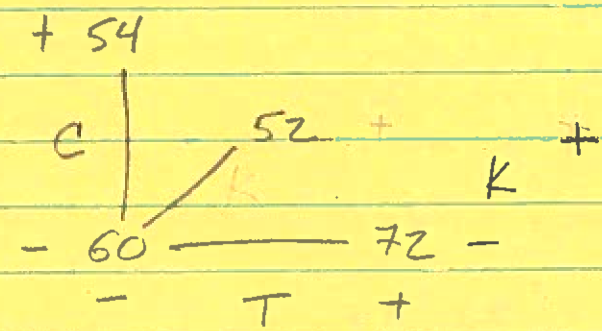


Figure 5.3. Display of the results from a pilot plant investigation employing a 2^3 factorial design to study the effects of **T** (temperature), **C** (concentration), and **K** (catalyst) on yield: (*a*) the 2^3 factorial with runs identified in standard order; (*b*) observed percent yields; (*c*) 12 treatment comparisons.

Optimizing w/ OFAT:

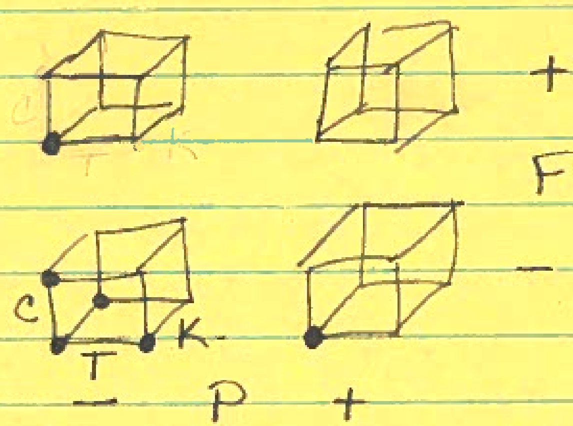


$\max_{OFAT} y = 72$ @ T+, K−, C−

$\max_{Fact} y = 83$ @ T+, K+, C−

This gets worse with more factors

add F: flow rate
P: pressure



• = runs in OFAT

Calculating Main Effects

OFAT

$ME(A) = \bar{y}(A+) - \text{control}, \quad \bar{y} = \text{mean}(y)$

For reactor problem

$ME(T) = \bar{y}(T+) - \text{control}$
$= 72 - 60 = 12$

$ME(C) = \bar{y}(C+) - \text{control}$
$= 54 - 60 = -6$

$ME(K) = \bar{y}(K+) - \text{control}$
$= 52 - 60 = -8$

Switching T from − to + increased yield by 12%.
Switching C from − to + decreased yield by 6%.

Issues

- Each effect depends heavily on the + case. Replication is essential!
- Findings are not Robust.
  - Data suggest we run at high T. When T=+, how should we set C & K? All our data for C+K were collected when T=−!

OFAT "recommends" $T = +$, $C = -$, $K = -$.

$y(T+, C-, K-) = 72.$

FACTORIAL DESIGN

$$ME(A) = \bar{y}(A+) - \bar{y}(A-)$$

Average over all other factor settings

For reactor problem

$$ME(T) = \bar{y}(T+) - \bar{y}(T-)$$
$$= \frac{72 + 68 + 83 + 80}{4} - \frac{60 + 54 + 52 + 45}{4} = 23$$

$$ME(C) = \bar{y}(C+) - \bar{y}(C-)$$
$$= \frac{54 + 68 + 45 + 80}{4} - \frac{60 + 72 + 52 + 83}{4} = -5$$

$$ME(K) = \bar{y}(K+) - \bar{y}(K-)$$
$$= \frac{52 + 83 + 45 + 80}{4} - \frac{60 + 72 + 54 + 68}{4} = 1.5$$

FACTOR	ME OFAT	ME FACTORIAL
T	12	23
C	-6	-5
K	-8	1.5

Factorial suggests $T+$, $C-$, $K+$, $y(T+, C-, K+) = 83$

Why did OFAT fail?

Maybe: we didn't explore enough
Actually: The variables <u>interact</u>

Remember: OFAT gave the wrong sign for ME(K); if it were correct, it would have suggested T+, C−, K+ like the factorial.

If T interacts with K, then ME(T) should depend on the level of K, and vice versa.

$$INT(TK) = \frac{ME(T|K+) - ME(T|K-)}{2}$$

conditional main effects (CME)

$$ME(T|K+) = ME(T) \text{ given } K=+$$

ME(T) conditioned on K = +

K = + in runs 5-8, so

$$ME(T|K+) = \frac{y_6 + y_8}{2} - \frac{y_5 + y_7}{2}$$

$$= \frac{83 + 80}{2} - \frac{52 + 45}{2}$$

$$= 81.5 - 48.5 = 33$$

Using runs 1-4 (K−)

$$ME(T|K-) = \frac{y_2 + y_4}{2} - \frac{y_1 + y_3}{2}$$

$$= \frac{72 + 68}{2} - \frac{60 + 54}{2}$$

$$= 70 - 57 = 13$$

$$\therefore\ INT(TK) = \frac{ME(T|K+) - ME(T|K-)}{2}$$

$$= \tfrac{1}{2}(33 - 13) = 10$$

The effect of moving T from − to + is greater when K = + (10% more yield)

<u>Symmetry</u>

$$INT(TK) = \tfrac{1}{2}\{ME(T|K+) - ME(T|K-)\}$$

$$= \tfrac{1}{4}(y_6 + y_8 - y_5 - y_7 - y_2 - y_4 + y_1 + y_3)$$

$$= \frac{y_1 + y_3 + y_6 + y_8}{4} - \frac{y_2 + y_4 + y_5 + y_7}{4}$$

$$INT(TK) = \tfrac{1}{2}\{ME(K|T+) - ME(K|T-)\}$$

$$= \frac{1}{2}\left\{\left(\frac{y_6 + y_8}{2} - \frac{y_2 + y_4}{2}\right) - \left(\frac{y_5 + y_7}{2} - \frac{y_1 + y_3}{2}\right)\right\}$$

$$= \frac{1}{4}(y_6 + y_8 - y_2 - y_4 - y_5 - y_7 + y_1 + y_3)$$

$$= \frac{y_1 + y_3 + y_6 + y_8}{4} - \frac{y_2 + y_4 + y_5 + y_7}{4}$$

Interactions as a contrast

y	Run	T	C	K	TK	TC	CK	TCK
60	1	−	−	−	+	+	+	−
72	2	+	−	−	−	−	+	+
54	3	−	+	−	+	−	−	+
68	4	+	+	−	−	+	−	−
52	5	−	−	+	−	+	−	+
83	6	+	−	+	+	−	−	−
45	7	−	+	+	−	−	+	−
80	8	+	+	+	+	+	+	+

$$INT(TC) = \bar{y}(TC+) - \bar{y}(TC-)$$
$$= \frac{y_1 + y_4 + y_5 + y_8}{4} - \frac{y_2 + y_3 + y_6 + y_7}{4}$$
$$= \frac{60+68+52+80}{4} - \frac{72+54+83+45}{4}$$
$$= \tfrac{1}{4}(260 - 254) = 1.5$$

$$INT(CK) = \bar{y}(CK+) - \bar{y}(CK-)$$
$$= \frac{y_1 + y_2 + y_7 + y_8}{4} - \frac{y_3 + y_4 + y_5 + y_6}{4}$$
$$= \frac{60+72+45+80}{4} - \frac{54+68+52+83}{4}$$
$$= \tfrac{1}{4}(257 - 257) = 0$$

Couldn't all 3 factors interact?

$$INT(TCK) = \tfrac{1}{2}\{INT(TC|K+) - INT(TC|K-)\}$$
$$= \tfrac{1}{2}\{INT(TK|C+) - Int(TK|C-)\}$$
$$= \tfrac{1}{2}\{INT(CK|T+) - Int(CK|T-)\}$$

or, using the TCK contrast

$$INT(TCK) = \tfrac{1}{4}(y_2 + y_3 + y_5 + y_8 - y_1 - y_4 - y_6 - y_7)$$

$$= \tfrac{1}{4}(72 + 54 + 52 + 80 - 60 - 68 - 83 - 45)$$

$$= 2/4 = 0.5$$

2-way interactions (2WI) are not common, so 3WI are even rarer.

	Contrast	Effect size
main effects	T	23
	C	-5
	K	1.5
2WI	TC	1.5
	TK	10
	CK	0
3WI	TCK	0.5

What matters most?

Definitely
T, C, TK

Probably
K, TC

Unlikely
CK, TCK

Table 5.6. A 2^3 Factorial Design to Study Effect of A (Length of Specimen), B (Amplitude of Load Cycle), and C (Load) in Investigation of Strength of Yarn

	Factors	Levels −	Levels +
A	Length, mm	250	350
B	Amplitude, mm	8	10
C	Load, g	40	50

Run Number	Factors A	Factors B	Factors C	Durance y
1	−	−	−	28
2	+	−	−	36
3	−	+	−	22
4	+	+	−	31
5	−	−	+	25
6	+	−	+	33
7	−	+	+	19
8	+	+	+	26

mean	27.5	
A	8	*
B	−6	*
C	−3.5	*
AB	0	
AC	−0.5	
BC	−0.5	
ABC	−0.5	

since only ME are significant, optimum should be at
A+, B−, C−

More difficult to decipher w/ interactions

Hidden Replication

Run	T	C	K
1	−	−	−
2	+	−	−
3	−	+	−
4	+	+	−
5	−	−	+
6	+	−	+
7	−	+	+
8	+	+	+

Run	T	C	K	
1	−	−	−	
3	−	+	−	2^2 in CK
5	−	−	+	
7	−	+	+	
2	+	−	−	
4	+	+	−	
6	+	−	+	2^2 in CK
8	+	+	+	

- Each effect (ME or INT) is a contrast involving all 8 runs.
- In OFAT, each effect contrasts 2 runs.
- Factorials have "hidden replication"

Why are factorials robust?

$$ME(K) = \bar{y}(K+) - \bar{y}(K-)$$

Run	T	C	K
5	−	−	+
6	+	−	+
7	−	+	+
8	+	+	+

Run	T	C	K
1	−	−	−
2	+	−	−
3	−	+	−
4	+	+	−

a factorial design for T & C!