Physics, chemistry, and chemical engineering knowledge & intuition

Bayesian network models to establish connections

Existing plant measurements

Patterns of likely causes & influences

Efficient experimental design to test combinations of causes

ANOVA & probabilistic models to eliminate irrelevant or uninteresting relationships

Process optimization (e.g. controllers, architecture, unit optimization, sequencing, and utilization)

Dynamical process modeling
**Scenario**
You have been called in as a consultant to find out how to optimize a client’s CSTR reactor system to both minimize product variation and also to maximize profit. After examining the whole dataset of 50 variables, you conclude that the most likely four variables for controlling profitability are the impeller type, motor speed for the mixer, control algorithm, and cooling water valve type. Your goal now is to design an experiment to systematically test the effect of each of these variables in the current reactor system.

These variables can take the following values:
- Impellers: model A, B, or C
- Motor speed for mixer: 300, 350, or 400 RPM
- Control algorithm: PID, PI, or P only
- Cooling water valve type: butterfly or globe

Each time you have to change the system setup, you have to stop much of the plant operation, so it means a significant profit loss.

**How should we design our experiment?**
Scenario

These variables can take the following values:
- Impellers: model A, B, or C
- Motor speed for mixer: 300, 350, or 400 RPM
- Control algorithm: PID, PI, or P only
- Cooling water valve type: butterfly or globe

Option 1: Factorial design to test all possible combinations

B, 300,PID, B  B, 300,PI, B  B, 300,P, B  B, 300,PI, G  B, 300,PID, G  B, 300,P, G
C, 300,PID, B  C, 300,PI, B  C, 300,P, B  C, 300,PI, G  C, 300,PID, G  C, 300,P, G
C, 400,PID, B  C, 400,PI, B  C, 400,P, B  C, 400,PI, G  C, 400,PID, G  C, 400,P, G

Total experiments = (3 impellers)(3 speeds)(3 controllers)(2 valves) = 54

Can we get similar information with fewer tests? How do we analyze these results?
Scenario

These variables can take the following values:
- Impellers: model A, B, or C
- Motor speed for mixer: 300, 350, or 400 RPM
- Control algorithm: PID, PI, or P only
- Cooling water valve type: butterfly or globe

Option 2: Taguchi Method of orthogonal arrays

Motivation: Instead of testing all possible combinations of variables, we can test all pairs of combinations in some more efficient way.

Key Feature:
Compare any pair of variables (P1, P2, P3, and P4) across all experiments and you will see that each combination is represented.

Example: L9 orthogonal array
Option 2: Taguchi Method of orthogonal arrays

Arrays can be quite complicated. Example: L36 array

Each pair of combinations is tested at least once
Factorial design: $3^{23} = 94,143,178,827$ experiments
Taguchi Method with L36 array: 36 experiments ($\sim 10^9 \times$ smaller)
Option 2: Taguchi Method of orthogonal arrays

Where do we these arrays come from?
1) Derive them
   • Small arrays you can figure out by hand using trial and error (the process is similar to solving a Sudoku)
   • Large arrays can be derived using deterministic algorithms (see http://home.att.net/~gsherwood/cover.htm for details)

2) Look them up
   • Controls wiki has a listing of some of the more common designs
   • Hundreds more designs can be looked up online on sites such as: http://www.research.att.com/~njas/oadir/index.html

How do we choose a design?
The key factors are the # of parameters and the number of levels (states) that each variable takes on.
Scenario

These variables can take the following values:
  Impellers: model A, B, or C
  Motor speed for mixer: 300, 350, or 400 RPM
  Control algorithm: PID, PI, or P only
  Cooling water valve type: butterfly or globe

Option 2: Taguchi Method of orthogonal arrays

# parameters: Impeller, speed, algorithm, valve = 4
# levels: 3 3 3 2 = ~3

<table>
<thead>
<tr>
<th>Experiment</th>
<th>Impeller</th>
<th>Motor Speed</th>
<th>Control</th>
<th>Valve</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>A</td>
<td>300</td>
<td>PID</td>
<td>BF</td>
</tr>
<tr>
<td>2</td>
<td>A</td>
<td>350</td>
<td>PI</td>
<td>G</td>
</tr>
<tr>
<td>3</td>
<td>A</td>
<td>400</td>
<td>P</td>
<td>G</td>
</tr>
<tr>
<td>4</td>
<td>B</td>
<td>300</td>
<td>PI</td>
<td>BF</td>
</tr>
<tr>
<td>5</td>
<td>B</td>
<td>350</td>
<td>P</td>
<td>BF</td>
</tr>
<tr>
<td>6</td>
<td>B</td>
<td>400</td>
<td>PID</td>
<td>G</td>
</tr>
<tr>
<td>7</td>
<td>C</td>
<td>300</td>
<td>P</td>
<td>G</td>
</tr>
<tr>
<td>8</td>
<td>C</td>
<td>350</td>
<td>PID</td>
<td>BF</td>
</tr>
<tr>
<td>9</td>
<td>C</td>
<td>400</td>
<td>PI</td>
<td>BF</td>
</tr>
</tbody>
</table>

No valve type 3, so this entry is filled at random in a balanced way
**Option 3: Random Design:** Surprisingly, randomly assigning experimental conditions will with high probability create a near optimal design.

- Choose the number of experiments to run (this can be tricky to do as it depends on how much signal recovery you want)
- Assign to each variable a state based on a uniform sample (e.g. if there are 3 states, then each is chosen with 0.33 probability)

Random designs tend to work poorly for small experiments (fewer than 50 variables), but work well for large systems.

For more information on these methods see the following resources:

http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=1614066
Scenario

These variables can take the following values:
  - Impellers: model A, B, or C
  - Motor speed for mixer: 300, 350, or 400 RPM
  - Control algorithm: PID, PI, or P only
  - Cooling water valve type: butterfly or globe

**When do we use which method?**

**Option 1: Factorial Design**
Small numbers of variables with few states (1 to 3)
Interactions between variables are strong and important
Every variable contributes significantly

**Option 2: Taguchi Method**
Intermediate numbers of variables (3 to 50)
Few interactions between variables
Only a few variables contributes significantly

**Option 3: Random Design**
Many variables (50+)
Few interactions between variables
Very few variables contributes significantly
Scenario

These variables can take the following values:
- Impellers: model A, B, or C
- Motor speed for mixer: 300, 350, or 400 RPM
- Control algorithm: PID, PI, or P only
- Cooling water valve type: butterfly or globe

Once we have a design, how do we analyze the data?

<table>
<thead>
<tr>
<th>Expt.</th>
<th>Impeller</th>
<th>Motor Speed</th>
<th>Control</th>
<th>Valve</th>
<th>Yield</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>A</td>
<td>300</td>
<td>PID</td>
<td>BF</td>
<td>16.1</td>
</tr>
<tr>
<td>2</td>
<td>A</td>
<td>350</td>
<td>PI</td>
<td>G</td>
<td>17.0</td>
</tr>
<tr>
<td>3</td>
<td>A</td>
<td>400</td>
<td>P</td>
<td>G</td>
<td>16.2</td>
</tr>
<tr>
<td>4</td>
<td>B</td>
<td>300</td>
<td>PI</td>
<td>BF</td>
<td>15.8</td>
</tr>
<tr>
<td>5</td>
<td>B</td>
<td>350</td>
<td>P</td>
<td>BF</td>
<td>17.9</td>
</tr>
<tr>
<td>6</td>
<td>B</td>
<td>400</td>
<td>PID</td>
<td>G</td>
<td>16.2</td>
</tr>
<tr>
<td>7</td>
<td>C</td>
<td>300</td>
<td>P</td>
<td>G</td>
<td>18.1</td>
</tr>
<tr>
<td>8</td>
<td>C</td>
<td>350</td>
<td>PID</td>
<td>BF</td>
<td>17.9</td>
</tr>
<tr>
<td>9</td>
<td>C</td>
<td>400</td>
<td>PI</td>
<td>BF</td>
<td>19.1</td>
</tr>
</tbody>
</table>

1) Plot the data and look at it
2) ANOVA
   1-way: effect of impeller
   2-way: effect of impeller and motor speed
Test multiple combinations
Scenario

These variables can take the following values:
- Impellers: model A, B, or C
- Motor speed for mixer: 300, 350, or 400 RPM
- Control algorithm: PID, PI, or P only
- Cooling water valve type: butterfly or globe

Once we have a design, how do we analyze the data?

<table>
<thead>
<tr>
<th>Expt.</th>
<th>Impeller</th>
<th>Motor Speed</th>
<th>Control</th>
<th>Valve</th>
<th>Yield</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>A</td>
<td>300</td>
<td>PID</td>
<td>BF</td>
<td>16.1</td>
</tr>
<tr>
<td>2</td>
<td>A</td>
<td>350</td>
<td>PI</td>
<td>G</td>
<td>17.0</td>
</tr>
<tr>
<td>3</td>
<td>A</td>
<td>400</td>
<td>P</td>
<td>G</td>
<td>16.2</td>
</tr>
<tr>
<td>4</td>
<td>B</td>
<td>300</td>
<td>PI</td>
<td>BF</td>
<td>15.8</td>
</tr>
<tr>
<td>5</td>
<td>B</td>
<td>350</td>
<td>P</td>
<td>BF</td>
<td>17.9</td>
</tr>
<tr>
<td>6</td>
<td>B</td>
<td>400</td>
<td>PID</td>
<td>G</td>
<td>16.2</td>
</tr>
<tr>
<td>7</td>
<td>C</td>
<td>300</td>
<td>P</td>
<td>G</td>
<td>18.1</td>
</tr>
<tr>
<td>8</td>
<td>C</td>
<td>350</td>
<td>PID</td>
<td>BF</td>
<td>17.9</td>
</tr>
<tr>
<td>9</td>
<td>C</td>
<td>400</td>
<td>PI</td>
<td>BF</td>
<td>19.1</td>
</tr>
</tbody>
</table>

3) Bin yield and perform Fisher’s exact test or Chi squared test to see if any effect is significant.
Field case study: Polyurethane quality control

Polyurethane manufacturing involves many steps, some of which involve poorly understood physics or chemistry.

Three dominant factors of product quality are:

1) Water content
2) Chlorofluorocarbon-11 (CFC-11) concentration
3) Catalyst type

Case modified from Lunnery, Sohelia R., and Joseph M. Sutej. "Optimizing a PU formulation by the Taguchi Method. (polyurethane quality control)." Plastics Engineering 46.n2 (Feb 1990): 23(5).
Field case study: Polyurethane quality control

<table>
<thead>
<tr>
<th>Factors and Levels</th>
<th># factors</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Polyol type</td>
<td>4</td>
<td>A,B,C, D</td>
</tr>
<tr>
<td>Catalyst package</td>
<td>5</td>
<td>1,2,3,4,5</td>
</tr>
<tr>
<td>Surfactant type</td>
<td>3</td>
<td>S1, S2, S3</td>
</tr>
<tr>
<td>Water, wt%</td>
<td>2</td>
<td>0.5, 1.5</td>
</tr>
<tr>
<td>CFC-11, wt%</td>
<td>2</td>
<td>25,35</td>
</tr>
<tr>
<td>Isocyanate type</td>
<td>2</td>
<td>11,12</td>
</tr>
</tbody>
</table>

*Case modified from* Lunnery, Sohelia R., and Joseph M. Sutej. "Optimizing a PU formulation by the Taguchi Method. (polyurethane quality control)." Plastics Engineering 46.n2 (Feb 1990): 23(5).
Field case study: Polyurethane quality control

Experiment design using a modified L16 array

A 3 S2 25 11  B 3 S3 25 11  Design modified from an L25 array to better account for the number of states of each variable.
B 1 S2 35 12  A 4 S3 35 12  Note not all pairs involving catalyst are tested--this is even sparser.
C 3 S1 35 12  D 3 S2 35 12
D 1 S1 25 11  C 4 S2 25 11
B 3 S1 25 12  A 3 S2 25 12
A 2 S1 35 11  B 5 S2 35 11
D 3 S2 35 11  C 3 S3 35 11
C 2 S2 25 12  D 5 S3 25 12

Case modified from Lunnery, Sohelia R., and Joseph M. Sutej. "Optimizing a PU formulation by the Taguchi Method. (polyurethane quality control)." Plastics Engineering 46.n2 (Feb 1990): 23(5).
Field case study: Polyurethane quality control

Experimental Procedure:
Reactivity profile and friability (subjective rating) were determined from hand-mix foams prepared in 1-gal paper cans. Free rise densities were measured on core samples of open blow foams. Height of rise at gel, final rise height, and flow ratio were determined in a flow tube.

Data Analysis
ANOVA to identify significant factors, followed by linear regression to identify optimal conditions

Case modified from Lunnery, Sohelia R., and Joseph M. Sutej. "Optimizing a PU formulation by the Taguchi Method. (polyurethane quality control)." Plastics Engineering 46.n2 (Feb 1990): 23(5).
Extreme Example: Sesame Seed Suffering

You have just produced 1000x 55 gallon drums of sesame oil for sale to your distributors. Just before you are to ship the oil, one of your employees remembers that one of the oil barrels was temporarily used to store insecticide and is almost surely contaminated. Unfortunately, all of the barrels look the same.

One barrel of sesame oil sells for $1000, while each assay for insecticide in food oil costs $1200 and takes 3 days. Tests for insecticide are extremely sensitive. **What do you do?**
**Extreme Example:** Sesame Seed Suffering

**Solution:** Extreme multiplexing. Like Taguchi methods but optimized for very sparse systems

Example solution w/ 8 barrels

Mix samples from each barrel and test mixtures

- Mix 1,2,3,4 --&gt; Sample A
- Mix 1,2,5,6 --&gt; Sample B
- Mix 1,3,5,7 --&gt; Sample C

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
</tr>
</thead>
<tbody>
<tr>
<td>A,B,C</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-,-,+</td>
<td>1</td>
<td></td>
<td></td>
<td>2</td>
<td>+,-,+</td>
<td>3</td>
<td>+,+,-</td>
<td></td>
</tr>
<tr>
<td>+,-,-</td>
<td>+,-,+</td>
<td></td>
<td>-,-,+</td>
<td></td>
<td>+,-,+</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>+,-,-</td>
<td>-,-,+</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Result: Using only 3 tests you can uniquely identify the poison barrel!

Is this enough tests?
Extreme Example: 
Sesame Seed Suffering

**Solution:** Extreme multiplexing. Like Taguchi methods but optimized for very sparse systems

**Solution w/ 1000 barrels**
Mix samples from each barrel and test mixtures
Experiments required=$\log_2(1000) \approx 10$

**Solution w/ 1,000,000 barrels**
Experiments required=$\log_2(1,000,000) \approx 20$

Optimal experiments can be extremely helpful!
Take Home Messages

• Efficient experimental design helps to optimize your process and determine factors that influence variability.

• Factorial designs are easy to construct, but can be impractically large.

• Taguchi and random designs often perform better depending on size and assumptions.