Design of Experiments:  
Part 2  

Dan Frey  
Asociate Professor of Mechanical Engineering and Engineering Systems
Plan for Today

• Adaptive experimentation
• Quasi-experimental design
• Philosophy of Science and Epistemology
One way of thinking of the great advances of the science of experimentation in this century is as the final demise of the “one factor at a time” method, although it should be said that there are still organizations which have never heard of factorial experimentation and use up many man hours wandering a crooked path.

My Observations of Industry

- Farming equipment company has reliability problems
- Large blocks of robustness experiments had been planned at outset of the design work
- More than 50% were not finished
- Reasons given
  - Unforseen changes
  - Resource pressure
  - Satisficing

“Well, in the third experiment, we found a solution that met all our needs, so we cancelled the rest of the experiments and moved on to other tasks...”
Minority Views on “One at a Time”

“…the factorial design has certain deficiencies … It devotes observations to exploring regions that may be of no interest…These deficiencies … suggest that an efficient design for the present purpose ought to be sequential; that is, ought to adjust the experimental program at each stage in light of the results of prior stages.”


“Some scientists do their experimental work in single steps. They hope to learn something from each run … they see and react to data more rapidly …If he has in fact found out a good deal by his methods, it must be true that the effects are at least three or four times his average random error per trial.”

Adaptive One Factor at a Time Experiments

If there is an apparent improvement, retain the change.
If the response gets worse, go back to the previous state.
Do an experiment

Change one factor

Stop after every factor has been changed exactly once.

If there is an apparent improvement, retain the change.
If the response gets worse, go back to the previous state.
Do an experiment
The First Step in aOFAT

\[ E(y(x_1^*, \tilde{x}_2, \ldots, \tilde{x}_n)) = E[\beta_1 x_1^*] + (n-1)E[\beta_{1j} x_1^* \tilde{x}_j] \]

\[ E[\beta_1 x_1^*] = \sqrt{\frac{2}{\pi}} \frac{\sigma_{ME}^2}{\sqrt{\sigma_{ME}^2 + (n-1)\sigma_{INT}^2 + \frac{1}{2}\sigma_{\varepsilon}^2}} \]

\[ E[\beta_{1j} x_1^* \tilde{x}_j] = \sqrt{\frac{2}{\pi}} \frac{\sigma_{INT}^2}{\sqrt{\sigma_{ME}^2 + (n-1)\sigma_{INT}^2 + \frac{1}{2}\sigma_{\varepsilon}^2}} \]

\[ \Pr(\beta_1 x_1^* > 0) = \frac{1}{2} + \frac{1}{\pi} \sin^{-1} \frac{\sigma_{ME}}{\sqrt{\sigma_{ME}^2 + (n-1)\sigma_{INT}^2 + \frac{1}{2}\sigma_{\varepsilon}^2}} \]
Performance after the First Step

\((n=7)\)
The Second Step in aOFAT

\[ E(y(x^*_1, x^*_2, x^*_3, \ldots, x^*_n)) = 2E[\beta_1 x^*_1] + 2(n-2)E[\beta_{1j} x^*_1] + E[\beta_{12} x^*_1 x^*_2] \]

\[ E[\beta_{12} x^*_1 x^*_2] = \sqrt{\frac{2}{\pi}} \frac{\sigma_{\text{INT}}^2}{\sqrt{\sigma_{\text{ME}}^2 + (n-1)\sigma_{\text{INT}}^2 + \frac{\sigma_\epsilon^2}{2}}} \]

\[ \Pr(\beta_{12} x^*_1 x^*_2 > 0) = \frac{1}{2} + \frac{1}{\pi} \tan^{-1} \frac{\sigma_{\text{INT}}}{\sqrt{\sigma_{\text{ME}}^2 + (n-2)\sigma_{\text{INT}}^2 + \frac{1}{2} \sigma_\epsilon^2}} \]
Performance after the Second Step

\( (n=7) \)

Expected improvement after the second variable is set in adaptive OFAT given a system with seven factors.
Probability of Exploiting the First Two-Factor Interaction \((n=7)\)

\[
\Pr(\beta_{12}x_1^*x_2^* > 0 \mid \beta_{12} > \beta_0) = \frac{1}{2} + \frac{1}{\pi} \tan^{-1} \frac{\sigma_{\text{INT}}}{\sqrt{\sigma_{\text{ME}}^2 + (n-2)\sigma_{\text{INT}}^2 + \frac{1}{2}\sigma_\epsilon^2}}
\]

\[
\frac{1}{\pi} \int_0^\infty \int_{-x_1}^{x_1} \left[ \frac{\epsilon}{\sqrt{2\sigma_{\text{INT}}}} \left( \frac{1}{\sqrt{2\sigma_{\text{INT}}}} \right)^{\frac{a}{2}} \frac{e^{-\frac{x_1^2}{2\sigma_{\text{INT}}^2}}}{\sigma_{\text{INT}}^2 + (n-2)\sigma_{\text{INT}}^2 + \frac{1}{2}\sigma_\epsilon^2} \right] dx_2 dx_1
\]

**Legend**

- Theorem 6
- \(\sigma_\epsilon / \sigma_{\text{ME}} = 0.1\)
- Simulation
- Theorem 6
- \(\sigma_\epsilon / \sigma_{\text{ME}} = 1\)
- Simulation
- Theorem 5
- \(\sigma_\epsilon / \sigma_{\text{ME}} = 10\)
- Simulation

- \(\Pr(\beta_{12}x_1^*x_2^* > 0)\)
- \(\Pr(\beta_{12}x_1^*x_2^* > 0 \mid \beta_{12} > \beta_0)\)
- \(\beta_{12}\)
- \(x_1\)
- \(x_2\)
- \(\sigma_{\text{INT}}\)
- \(\sigma_{\text{ME}}\)
- \(\sigma_\epsilon\)
- \(\pi\)
- \(\tan^{-1}\)
- \(\sqrt{\cdot}\)
- \(\int\)
- \(\epsilon\)
- \(\sigma_{\text{INT}}\)
- \(\sigma_{\text{ME}}\)
- \(\sigma_\epsilon\)
- \(\frac{1}{\pi}\)
- \(x_1\)
- \(x_2\)
- \(\text{Simulation}\)
- \(\text{Theorem 6}\)
- \(\text{Theorem 5}\)
- \(\text{Theorem 6}\)
- \(\text{Theorem 5}\)
- \(\text{Simulation}\)
- \(\text{Simulation}\)
Performance after Multiple Steps

\[
\Pr(\beta_{ij}x_i^*x_j^* > 0) \geq \Pr(\beta_{12}x_1^*x_2^* > 0)
\]
Final Outcome (n=7)

Adaptive OFAT

Resolution III Design
Final Outcome \((n=7)\)

Adaptive OFAT

Resolution III Design
## Electric Airplane Experiment

<table>
<thead>
<tr>
<th>Factor</th>
<th>Description</th>
<th>Level 1</th>
<th>Level 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>Propeller diameter</td>
<td>7 in.</td>
<td>8 in.</td>
</tr>
<tr>
<td>B</td>
<td>Propeller pitch</td>
<td>4 in.</td>
<td>5 in.</td>
</tr>
<tr>
<td>C</td>
<td>Gear ratio</td>
<td>1:1</td>
<td>1:1.85</td>
</tr>
<tr>
<td>D</td>
<td>Wing Area</td>
<td>450 in(^2)</td>
<td>600 in(^2)</td>
</tr>
<tr>
<td>E</td>
<td>Cells in battery</td>
<td>7</td>
<td>8</td>
</tr>
<tr>
<td>F</td>
<td>Motor Type</td>
<td>SP400 7.2V</td>
<td>SP480 7.2V</td>
</tr>
<tr>
<td>G</td>
<td>Number of motors</td>
<td>1</td>
<td>2</td>
</tr>
</tbody>
</table>

From Unified Engineering 16.01-16.04
Electric Airplane – Active Effects

If there is are two motors, the increase in wing area is advantageous.

If the motor is geared down, the increase propeller diameter is advantageous.

Response = Maximum flight time

<table>
<thead>
<tr>
<th>Term</th>
<th>Coefficient</th>
</tr>
</thead>
<tbody>
<tr>
<td>C</td>
<td>9.71</td>
</tr>
<tr>
<td>G</td>
<td>5.10</td>
</tr>
<tr>
<td>E</td>
<td>3.58</td>
</tr>
<tr>
<td>F</td>
<td>-3.24</td>
</tr>
<tr>
<td>D*G</td>
<td>1.91</td>
</tr>
<tr>
<td>A*C</td>
<td>1.43</td>
</tr>
<tr>
<td>C<em>F</em>G</td>
<td>-1.13</td>
</tr>
<tr>
<td>E*G</td>
<td>0.90</td>
</tr>
<tr>
<td>B*C</td>
<td>0.83</td>
</tr>
<tr>
<td>D<em>E</em>G</td>
<td>0.83</td>
</tr>
<tr>
<td>C<em>D</em>E*F</td>
<td>0.79</td>
</tr>
<tr>
<td>B</td>
<td>-0.79</td>
</tr>
<tr>
<td>B*G</td>
<td>0.38</td>
</tr>
<tr>
<td>A*F</td>
<td>-0.35</td>
</tr>
</tbody>
</table>
Electric Airplane Results

- Maximum flight time among 128 designs
- Average flight time among 128 designs
- Expected value with OFAT
- Range from 10th to 90th percentile
Adaptive “One Factor at a Time” for Robust Design

Run a resolution III on noise factors

Again, run a resolution III on noise factors. If there is an improvement, in transmitted variance, retain the change.

Change one factor

If the response gets worse, go back to the previous state.

Stop after you’ve changed every factor once.

A Manufacturing Case Study

- Sheet metal spinning
- 6 control factors (number of passes of the tool, etc.)
- 3 noise factors (material properties, etc.)
- Goal = more uniform geometry

Kunert, J., et. al., 2004, “An experiment to compare the combined array and the product array for robust parameter design,” accepted to *J. of Quality Technology*. 

A Manufacturing Case Study

- aOFAT worked better if experimental error not too high
- Especially true if an informed starting point was used

Results Across Four Case studies

<table>
<thead>
<tr>
<th>Method used</th>
<th>Fractional array $\times 2^{k-p}_{III}$</th>
<th>aOFAT $\times 2^{k-p}_{III}$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>informed</td>
<td>random</td>
</tr>
<tr>
<td>sheet metal spinning</td>
<td>Low $\varepsilon$</td>
<td>51%</td>
</tr>
<tr>
<td></td>
<td>High $\varepsilon$</td>
<td>36%</td>
</tr>
<tr>
<td>op amp</td>
<td>Low $\varepsilon$</td>
<td>99%</td>
</tr>
<tr>
<td></td>
<td>High $\varepsilon$</td>
<td>98%</td>
</tr>
<tr>
<td>paper airplane</td>
<td>Low $\varepsilon$</td>
<td>43%</td>
</tr>
<tr>
<td></td>
<td>High $\varepsilon$</td>
<td>41%</td>
</tr>
<tr>
<td>freight transport</td>
<td>Low $\varepsilon$</td>
<td>94%</td>
</tr>
<tr>
<td></td>
<td>High $\varepsilon$</td>
<td>88%</td>
</tr>
<tr>
<td>Mean of four cases</td>
<td>Low $\varepsilon$</td>
<td>74%</td>
</tr>
<tr>
<td></td>
<td>High $\varepsilon$</td>
<td>66%</td>
</tr>
<tr>
<td>Range of four cases</td>
<td>Low $\varepsilon$</td>
<td>43% to 99%</td>
</tr>
<tr>
<td></td>
<td>High $\varepsilon$</td>
<td>36% to 88%</td>
</tr>
</tbody>
</table>

Courtesy of ASME. Used with permission.
Conclusions: Adaptive Experimentation

• If the goal is maximum improvement rather than maximum precision in estimation
• And experimental error is not too large
• And sequential experiments are possible
• Then adaptive experimentation provides significant advantages over factorial plans
• Mostly because it exploits two-factor interactions, especially the largest ones
• Proven to be effective for robust design
Plan for Today

• Adaptive experimentation

Quasi-experimental design

• Philosophy of Science and Epistemology
Quasi-Experimental Design

• Treatments do not meet fully the criteria of an experiment
  – Not actually applied by the experimenter, but occurred "naturally"
  – OR not randomized
  – OR no control group

• Primary techniques
  – Comparison group design
  – Interrupted time series

• Key issue – entertain seriously the alternative hypotheses
A Comparison Group Study

• One area of the country chosen (Boston)
• Three groups of homes known to have been weatherized in a certain year
• Control = randomly selected homes not among those known to be weatherized

Figure 1. Average Home Primary Heating Fuel Bills per Household

An Interrupted Time Series Study

- Small town puts a smoking ban in place
- Reduced incidence of admissions for myocardial infarction observed

An Educational Study

• Teaching Method A is used by the majority of high school teachers in a district.

• The option of adopting method B is offered and some percentage accept

• Training is provided and some extra pay

• Method B results in X% better pre-test to post-test improvements in raw score
Volunteer Effects

• Ask people to volunteer for a new activity
• Those who volunteer (as compared with those who don't) are on average
  – Higher IQ
  – Younger
  – More approval seeking
  – Different in psychological adjustment
    • If a behavioural study, better adjusted
    • If a medical study, mal-adjusted

# Ceiling Effects

<table>
<thead>
<tr>
<th>Method A (non-volunteer)</th>
<th>Method B</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pre-test score</td>
<td>Pre-test score</td>
</tr>
<tr>
<td></td>
<td>Post-test score</td>
</tr>
<tr>
<td>80</td>
<td>40</td>
</tr>
<tr>
<td></td>
<td>70</td>
</tr>
</tbody>
</table>

If the paper reports "the pre-post- test gains for method B were significantly higher" what would you say about that?
Plan for Today

• Adaptive experimentation
• Quasi-experimental design

Philosophy of Science and Epistemology
Concept Question

• Each card has a letters on one side and a number on the other
• Hypothesis-*if* a card has a D on one side it *must* have a 3 on the other side
• You are a scientist investigating this hypothesis
• You are allowed to turn over any *two* cards
• Which would you choose to turn over?

1) D&F  2) D&3  3) D&7  4) F&3  5) F&7  6) 3&7
David Hume:
The Problem of Induction

• "We have no other notion of cause and effect, but that of certain objects, which have been always conjoin'd together, and which in all past instances have been found inseparable. We cannot penetrate into the reason of the conjunction. We only observe the thing itself, and always find that from the constant conjunction the objects acquire a union in the imagination."

Nelson Goodman: The New Riddle of Induction

- Some regularities in the world establish habits, some do not
- Proposes the color "grue"
  - Applies to all things examined before a certain time $t$ just in case they are green
  - But also to other things just in case they are blue and not examined before time $t$
- But there is a virtuous circle that makes induction work

Ernst Mach: Phenomenolism

- *Analysis of Sensations* (1885)
- Postulates "elements" such as individual sounds, temperatures, pressures, spaces, times, and colors
- All material things including our own bodies are nothing but complexes of elements that have been constructed by the human mind
- Material bodies do not produce sensations, but rather complexes of sensations are associated together by the human mind to produce material bodies
Hegel's Dialectic

- A dialectic of existence
  - First, existence must be posited as pure Being (Sein)
  - But pure Being, upon examination, is found to be indistinguishable from Nothing (Nichts)
  - Being and Nothing are united as Becoming

Logical Positivism

• “…it began in the 1920’s and flourished for about twenty or thirty years …they were convinced that a genuine contingent assertion about the world must be verifiable through experience and observation.”

  Cambridge Dictionary of Philosophy

• Analytic / synthetic dichotomy. Analytic truths are true (or false) by virtue of some rules of language (including math).
Popper: Falsificationism

- The criterion of demarcation of empirical science from pseudo science and metaphysics is falsifiability.
- The strength of a theory can be measured by the breadth of experimental results that it precludes.

Sir Karl Popper, *Logik der Forschung*
Suhs Independence "Axiom"

Maintain independence of the functional requirements

\[
\{\text{FR}\} = [A]\{\text{DP}\}
\]

where

\[
A_{i,j} = \frac{\partial \text{FR}_i}{\partial \text{DP}_j}
\]

\[
\begin{bmatrix}
X & 0 & 0 \\
0 & X & 0 \\
0 & 0 & X
\end{bmatrix}
\]

uncoupled

\[
\begin{bmatrix}
X & 0 & 0 \\
X & X & 0 \\
X & X & X
\end{bmatrix}
\]

decoupled

\[
\begin{bmatrix}
X & X & X \\
X & X & X \\
X & X & X
\end{bmatrix}
\]

coupled

Acceptable

Avoid

Kuhn: Scientific Revolutions

- Paradigm = a set of scientific and metaphysical beliefs that make up a theoretical framework within which scientific theories can be tested, evaluated, and revised.
- “Normal science” = refinement within a paradigm
- “Revolution” = older paradigm overthrown

Lakatos

• Resolves Popper and Kuhn
• Research programs
  – Progressive programs generate bold predictions and useful new work although they may have some counterevidence against them
  – Degenerate programs seek to defend their theory against all evidence and may even do so successfully, but make no useful predictions
Concept Question

• Each card has a letters on one side and a number on the other
• Hypothesis- *if* a card has a D on one side it *must* have a 3 on the other side
• You are a scientist investigating this hypothesis
• You are allowed to turn over any *two* cards
• Which would you choose to turn over?

1) D&F  2) D&3  3) D&7  4) F&3  5) F&7  6) 3&7
Overview Research

Adaptive Experimentation and Robust Design

\[ \Pr(\beta_1 x_i^* x_i^* > 0 | \beta_i > \beta_j) > \frac{1}{\pi} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \left( \frac{1}{\sqrt{2 \pi \sigma_{\text{INT}}^2}} \right)^2 \exp \left( -\frac{(x_i - \mu_0)^2}{2 \sigma_{\text{INT}}^2} \right) \left( -\frac{(x_i - \mu_1)^2}{2 \sigma_{\text{INT}}^2} \right) dx_i dx_i \]

Complex Systems

Methodology Validation

Outreach to K-12

PBS show "Design Squad"
Next Steps

• Wednesday 2 May
  – Design of Computer Experiments
• Friday 4 May
  – Exam review
• Monday 7 May – Frey at NSF
• Wednesday 9 May – Exam #2