Tutorial on "Real-World DOE Problems" — Tom Donnelly, 15 June 2009, 77th MORSS

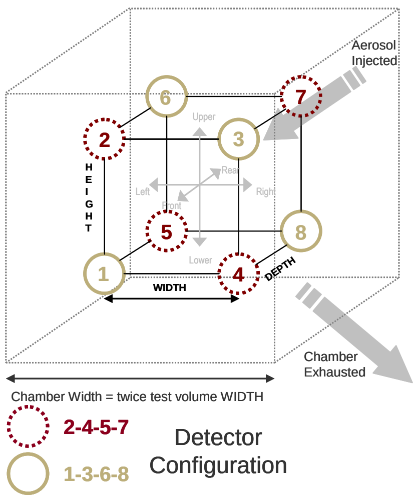
This tutorial is intended to complement the introductory Design of Experiments (DOE) tutorial. You may not learn as much "how to," but will learn to recognize real-world problems and when you need to get help from a knowledge source or someone with experience. It's far better to go to a DOE subject matter expert (SME) before you experiment then after.

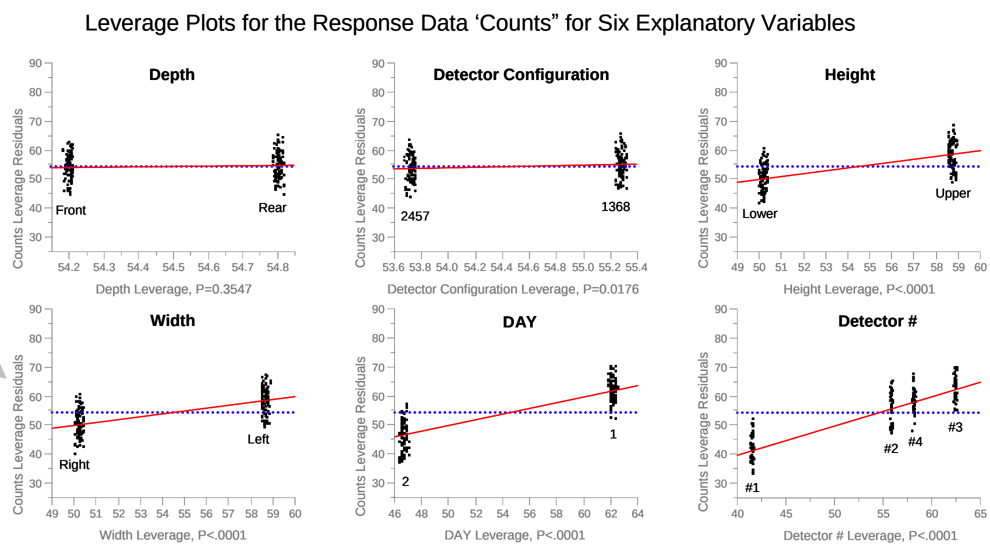
<u>Why use Design of Experiments (DOE) methods?</u>
It is the most cost effective way to get quick answers to multivariable problems. Another way to think about it is for your existing budget you can solve more and/or bigger problems.

<u>Why is using DOE methods important?</u>
DOE is one of the more powerful tools we can use to quickly and efficiently develop and optimize the multivariable technologies needed to best equip and protect our warfighters.

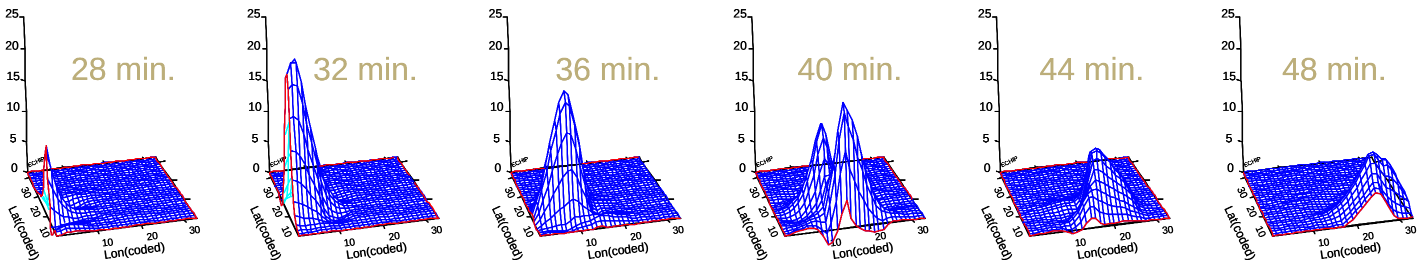
<u>Some easy things to do</u>:

<u>Plot the data:</u> Plot ALL the data when possible. Leverage plots are handy. Present plots in small multiples. See books by Edward Tufte (www.edwardtufte.com). Tufte's first grand principle is "Show Comparisons." Here are two examples: The first example shows plots of all the observed data from a study of aerosol particle counts in a chamber for six explanatory variables.





The second example shows aerosol concentration versus location variables for 6 time steps from analysis of simulation data using a Kriging regression technique.



<u>Checkpoints</u>, checkpoints, checkpoints: – If you can't predict your response within a certain tolerance, how good can your model be? Consider taking checkpoints inside and outside design space, at optima, where unusual behavior is observed, that support the next higher model, the boss' suggestion, during the design… Compare observed value and model prediction ± limits. Compare residual SD (model error) and checkpoint RMS (checkpoint error) – they should be similar.

<u>Replicate at Least 5 Trials</u> – and 10 would be better. Then you can conduct a lack-of-fit test comparing model error (Residual SD) to pure error (Replicate SD). The actual F-test used compares the variances (squares of the SDs) and takes into account the number of degrees of freedom in each variance estimate.

Tom Donnelly, PhD • SAS Institute Inc.
302.737.1650 • tom.donnelly@jmp.com

Before running your real-world design, do an analysis with fictitious data to check the correlation among the factors. Also be sure to run a feasibility analysis of the trials and consider doing the hardest one first.

Historical (Hysterical?) Data – Existing data may hold some useful information, certainly about the process variability, but way too often it doesn't broadly cover the variable space of interest – lots of data, but too much of the same thing. Generally better off starting fresh.

RESOURCES: By no means is this an exhaustive listing!

The first DOE book:
Fisher, R. A. (1935), *The Design of Experiments*, Oliver and Boyd

Textbooks this Century:
1. Box, G. E. P., Hunter, W. G., and Hunter, J. S. (2005), *Statistics for Experimenters*, 2nd ed., Wiley, New York
   - The standard – classic 1978 text recently revised
2. Wu, C. F. J. and Hamada, M. (2000), *Experiments, Planning, Analysis and Parameter Design Optimization*, Wiley, New York
   - Both classic approaches and orthogonal arrays & orthogonal main effects plans
3. Montgomery, D. C. (2009), *Design and Analysis of Experiments*, 7th ed., Wiley, New York
   - Popular text, solution book available, examples illustrated with JMP® software.

More Specialized Texts – Optimal Design, Mixtures, Response Surfaces:
1. Atkinson, A. C. and Donev, A. N. (1992), *Optimum Experimental Designs*, Clarendon Press, Oxford
2. Cornell, J. A. (2002), *Experiments with Mixtures, Designs, Models and the Analysis of Mixture Data*, 2nd ed. Wiley, New York
3. Khuri, A. I. and Cornell, J. A. (1996), *Response Surfaces, Designs and Analyses*, 2nd ed., Marcel Dekker, New York
4. Box, G. E. P. and Draper, N. A. (2007), *Response Surfaces, Mixtures and Ridge Analysis*, 2nd ed., Wiley, New York
5. Myers, R. H. and Montgomery, D. C. (2002), *Response Surface Methodology: Process and Product Optimization Using Designed Experiments*, 2nd ed., Wiley, New York

Texts Specifically on DOE for Computer Experiments:
1. Kleijnen, J. P. C. (2008), *DASE: design and analysis of simulation experiments*. Springer, New York.
2. Santner, T. J., Williams, B. J., and Notz, W. I. (2003), *The Design and Analysis of Computer Experiments*, Springer, New York
3. Fang, K. T., Li, R. Z., and Sudjianto, A. (2005), *Design and Modeling for Computer Experiments*, Chapman & Hall/CRC Press, New York

One more book:
Good, P. I. and Hardin, J. W. (2006), *Common Errors in Statistics (and How to Avoid Them)*, Wiley, New York

Software & Code:
1. DOE specific applications:
   a. Design-Ease/Design-Expert www.statease.com/
   b. ECHIP
2. General statistical analysis applications with strong DOE modules:
   a. JMP www.jmp.com/
   b. Minitab www.minitab.com/
3. The "R" statistical open source project www.r-project.org/,
4. Matlab www.mathworks.com/ specialized modules for kriging and metamodeling:
   a. DACE www2.imm.dtu.dk/~hbn/dace/
   b. SUMO www.sumo.intec.ugent.be/
5. The SEED Center for Data Farming at Naval Postgraduate School http://harvest.nps.edu/ Code/Modules to create
   a. Nearly Orthogonal Latin Hypercubes (NOLH) designs for computer simulation experiments
   b. Resolution V, Fractional-Factorial designs for many factors
6. Homepage of Prof. Hongquan Xu of UCLA, Statistics Dept.: Source code for algebraic orthogonal array generator www.stat.ucla.edu/~hqxu/nsf/

Websites for Orthogonal Arrays:
1. http://www.research.att.com/~njas/oadir/index.html Library of Orthogonal Arrays maintained by Neil J.A. Sloane
2. http://support.sas.com/techsup/technote/ts723.html Library of Orthogonal Arrays maintained by Warren F. Kuhfield

Tom Donnelly, PhD • SAS Institute Inc.
302.737.1650 • tom.donnelly@jmp.com

20 (some) Questions I Like to Ask at the Start of DOE Discussions:

1. What is the goal of the experimentation?
2. How do you measure success?
3. What response variables do you measure?
4. What are all the control factors that may affect these responses?
5. Over what ranges does it make sense to operate these variables?
6. Do any combinations of variable settings cause problems? (Safety? Cost? Breaks the equipment? Impossible to achieve?)
7. Do you currently run control samples for this process?
8. If you do exactly the same process on separate days do you ever get obviously/surprisingly different results?
9. How big is the variability (What is the standard deviation?) for each response?
10. Do you have past records of replicated trials for each response?
11. Are the replicate trials close together or spread out over time?
12. How tiny of a difference for each response is considered practically important?
13. Do you think we are looking for tiny differences in big variability (hard to do because lots of replication is needed) or big differences in small variability (easy)?
14. If more than one response needs to be characterized for your process, what is their relative importance?
15. Are you interested in identifying the best trade-off in performance of several responses?
16. Are you more interested in identifying important control factors or in ending up with a model that can predict your responses?
17. How many trials can be run in a day?
18. Are there any hard-to-change factors?
19. How many devices do you have of each type?
20. How hard is it to come back at a later time to run checkpoint trials?
21. What is your budget?
22. What is your deadline?

Summary Part 1 – Introduction and Response Surface Methods:
Five steps to optimize a process:
1. Think about process – see list of 20 (some) questions above
2. Do some work – not just any work will do - design is the set of trials run to support the proposed model
3. Analyze data – model fits? – even if it doesn't look at the pictures of the process
4. Optimize Process – look at the pictures – minimum? maximum? target?
5. Verify results – checkpoints, checkpoints, checkpoints

Expensive or time consuming trials? Use a sequence of designs to successively increase the complexity of the model supported.

Summary Part 2 – Multiple Response Optimization:
Determine the best tradeoff in performance among several responses. This is a highly interactive process. Always verify the optimum with checkpoints.

Summary Part 3 – Algorithmic Design:
Use a textbook design when ever you can. When they don't fit your problem use an algorithmic solution (software).

Algorithmic designs can be created for all these problems:
1. special models,
2. combinations of any or all these types of variables:
    a. continuous (quantitative), - finely adjustable like *temperature*, *speed* or *force*
    b. categorical (qualitative), - comes in types like material = *wood*, *plastic* or *metal* with mixed numbers of levels (3 materials, 4 machines, and 5 operators)
    c. mixture (formulation) – *blending* ingredients - process depends more on the *proportions* than on the amount
    d. and blocking – a variable for which there "shouldn't be" a causal effect – *day*, *lot*, *batch*, *tray*
3. constrained regions (constraints),
4. adding on to existing trials (augmentation),
5. repairing broken designs (both constraints and augmentation),

Summary Part 4 – Mixtures & Combining Variable Types:
Examples with 4 and 5 mixture components with and without algebraic inequality constraints were demonstrated.
One complex design example was demonstrated with 10 variables - 6 mixture, 2 continuous, 1 categorical and 1 blocking - with additional constraints.

Issues specific to some mixture designs:
1. Absence (0%) vs. presence (0.1%) can have greater effect than change from 1% to 10% for some components – catalysts, dopants, etc.
2. "Additive" (in a mathematical sense) mixture component doesn't take part in the chemistry – filler, binder, colorant, diluent
3. Some component(s) are held at a constant value forcing balance to sum to less than 1.
4. Trace quantities act like process variables (0.0001 to 0.0002).

Summary Part 5 – "One shot" Designs – Special Orthogonal Arrays:
Only use if you really only get "one shot" at it!
Depends strongly on the assumptions
1. *Factor sparsity* – only a few variables are active in a factorial experiment
2. *Effect heredity* – significant interactions only appear among these active factors

A 10-variable example is shown that in 27 trials yields a response surface model in the 4 most important factors.
See these references:
1. Cheng, S.W. & Wu, C.F. J. "Factor Screening and Response Surface Exploration(with discussion)." *Statistica Sinica*, 11 (2001), No. 3, 553-604.
2. Xu, H., Cheng, S.W. & Wu, C. F. J. "Optimal Projective Three-Level Designs for Factor Screening and Interaction Detection." *Technometrics*, 46 (2004), 280-292.

Summary Part 6 – Transformations:
They're free in the sense that you don't need more data! Can help data meet regression assumption of being normally distributed with constant variance with the benefits of often eliminating lack-of-fit and preventing model predictions from being nonsensical (e.g. negative # of defects/unit area, yield > 100%, etc.) - Especially useful when data values run up against a boundary. The most frequently used transformations are:
For data bounded on the low side (e.g. # defects, resistivity, hardness, etc.):

$Y = \text{Log}(y)$ – used when values range over several orders of magnitude

$Y = y^{1/2}$ – used with counting data

For data bounded on two sides (e.g. percentage range of 0 -100%, or a range of 1 = worst to 9 = best)

$Y = 2*\text{arcsine}(y^{1/2})$ – used for values scaled between 0 and 1

To scale a bounded range (low, high) to the unit range (0, 1) use $Y = (y – \text{low})/(\text{high} – \text{low})$

For Pass/Fail (binary) data, IF all trials have the same number of attempts, then you can use the transformation $Y = 2*\text{arcsine}(y^{1/2})$, but a better tool is logistic regression. First time using logistic regression I suggest you enlist help of an SME.
A great reference: *Plots, Transformations and Regression*, A.C. Atkinson, (1984), Oxford University Press

Summary Part 7 – DOE for Computer Experiments:
Currently becoming a hot topic since so many DoD programs require M&S.
DOE is used to metamodel the long-running simulation model.
Both traditional (factorial type) designs and "space-filling" designs are discussed.
Traditional designs were run in a sequential fashion until adequate accuracy was obtained for one example.
Space filling designs can be analyzed using kriging methods when the response is – non-stochastic (non-random) and when all variables are continuous although these limitations are actively being pursued in academia today.
Some good references on these topics include:
1. "Blind Kriging: A New Method for Developing Metamodels," Joseph, V.R., Hung, Y., and Sudjianto, A., *ASME Journal of Mechanical Design*, 130, 031102-1-8, 2008
2. "Gaussian Process Models for Computer Experiments With Qualitative and Quantitative Factors," Qian, P.Z.G., Wu, H., and Wu, C.F.J., to appear in *Technometrics*, 2008
3. "Building surrogate models with detailed and approximate simulations," Qian, Z., Seepersad, C., Joseph, R., Wu, C. F. J., and Allen, J., *Proceedings of the IDETC'04 ASME 2004 Design Engineering Technical Conferences and Computers and Information in Engineering Conference*

Summary Part 8 – Robust Product Design:
Two main approaches:
1. Robust Tolerances – exploit the "flatter" regions of the response vs. control variable model
    a. Assumes model accurately reflects the process and "transmits" the variation
    b. Assumes tolerance accurately reflect the variation in the control variables
2. Simultaneously minimize the Variance while Targeting the Mean
    a. Takes more data to model the variance than the mean

"The purpose of models is not to fit the data but to sharpen the questions." – Samuel Karlin

Tom Donnelly, PhD • SAS Institute Inc.
302.737.1650 • tom.donnelly@jmp.com