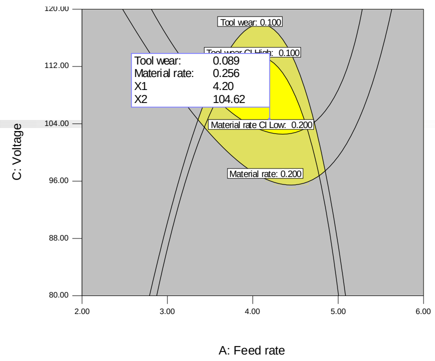
4th European DoE User Meeting

Stat-Ease statistics made easy

27.06. - 28.06.2012
--
Vienna, Austria

STATCON

Vielen Dank Heidi & Claudia



Modern Design of Experiments (DOE)— 75 Years of Advancements in Multifactor Test Methods

Mark J. Anderson, PE, CQE (Speaker)
Patrick J. Whitcomb, PE
Stat-Ease, Inc., Minneapolis, MN
mark@statease.com 612-746-2032

Milestones in Development of Modern DOE

<300 BC through 19th Century: Aristotle, Bacon, Edison – Scientific Method of One Factor at a Time (OFAT)

1886: Galton – Regression

1908: Gossett – Simple comparative testing

1926: Fisher – Multiple comparative and multifactor testing

1951: Box – Response Surface Methods (RSM)

1959: Box & Lucas – Optimal design

1960s: DuPont – SCO (Screening, Characterization, Optimization) strategy

1964: Box & Cox – Data transformations

1964: Scheffe – Mixture design

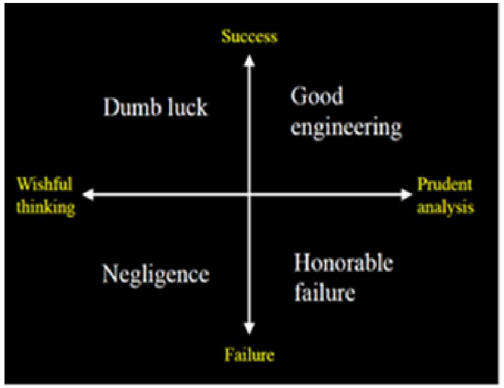
1980s: Taguchi – Robust design

2002: Whitcomb-Oehlert – Minimum-Run Designs

Current: Regulated Industries – Quality by Design (QbD)

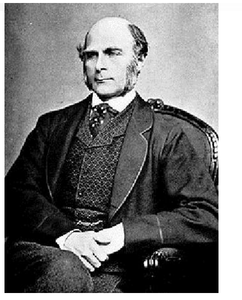
Before Statistical Methods
How Industrial Experimenters Succeeded

1. Scientific method: Commonly attributed to Francis Bacon in the 17th century, stemming from Aristotle in mid-300s BC.
2. Persistence: Edison's 1% inspiration and 99% perspiration.
3. Good engineering: Edison's protégé Charles Steinmetz (pictured with him above) once charged $1000 to GE for knowing which part to investigate on an electrical device, $1 for the chalk mark and $999 for knowing where to put it.
4. "Dumb luck"!



Source: "Beyond Probability, A pragmatic approach to uncertainty quantification in engineering" Scott Ferson, NASA Statistical Engineering Symposium, Williamsburg, Virginia, 4 May 2011

Regression of Happenstance Data *(1/2)*



Regression analysis, invented in the late 19th century by Francis Galton (pictured),* connects the responses (Y's) to the input factors (X's) via mathematical models of the form: $\hat{Y} = \beta_0 + \beta_1 X_1 + \beta_2 X_2 \ldots + \beta_k X_k + \varepsilon$

where k is the number of factors and ε represents error.

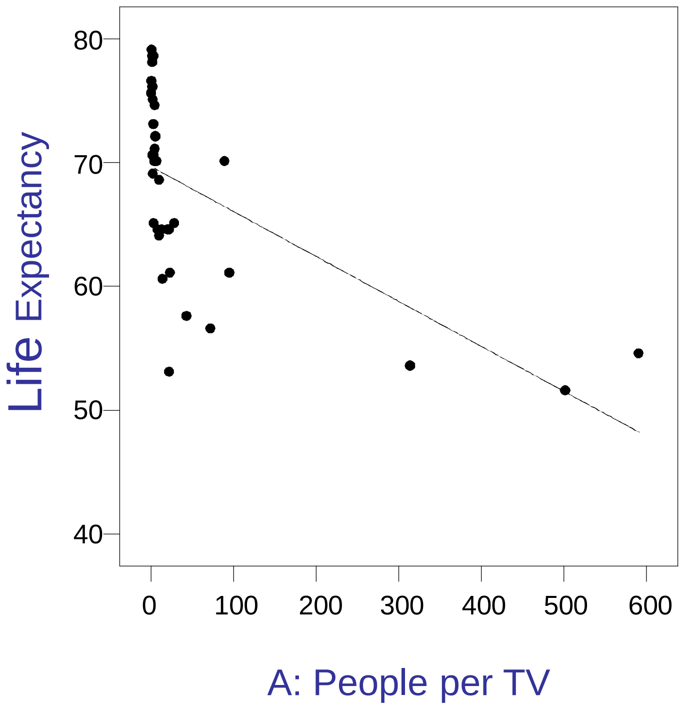
*"Regression towards mediocrity in hereditary stature". *The Journal of the Anthropological Institute of Great Britain and Ireland* (1886). 15: 246–263

"Engineers are quite comfortable these days - in fact, far too comfortable – with results from the blackest of black boxes: neural nets, genetic algorithms, data mining, and the like."

- Russell Lenth (Professor of Statistics, University of Iowa)

Regression of Happenstance Data *(2/2)*

A Cal Poly stats prof observed* that life expectancy in various countries varies with the number of people per television (TV). This solves our problems replacing old CRT and low-def units: Ship them to the Third World so these poor TV-deprived people can live longer! ;) (Wink.)



*Allan Rossman , "Televisions, Physicians, and Life Expectancy." *Journal of Statistics Education* 2, no. 2 (1994).

PS. Breaking News: When the TVs arrive, do not turn them on!

22 MINUTES OF LOST LIFE FOR EVERY HOUR OF TV WATCHED

"Could TV Shorten Your Lifespan?" 8/19/11 www.huffingtonpost.com

Simple Comparative Experiments

A century ago William Sealy Gossett, a chemist at Guiness Brewery, developed a statistical method called the “t-test” to determine when the yeast content of a particular batch of beer differed significantly from the brewery's standard.*

This is a simple comparative experiment on one factor at a time (OFAT). It is still widely used of sensory and other evaluations.

*(Published in 1908 under the pseudonym “Student”.)

A Very Small Dose of Stat Detail

One Factor Comparison via t-Tests

Legal judgment: Innocent until proven guilty.

Hypothesis test: Same until proven different.

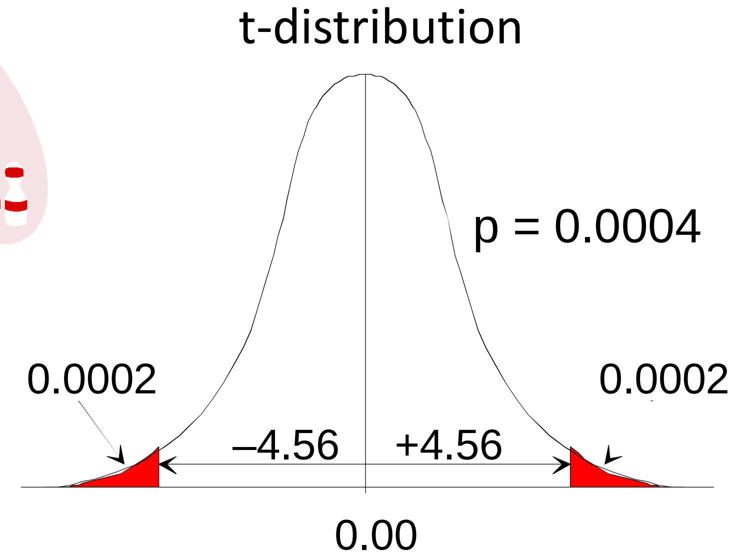
$$H_0(\text{"null"}):\ \mu_1 = \mu_2\ \ \text{(samples from same population)}$$

$$H_1(\text{"alternative"}):\ \mu_1 \neq \mu_2\ \ \text{(samples from different populations)}$$

$$t = \frac{\bar{Y}_1 - \bar{Y}_2}{s_{\bar{Y}_1 - \bar{Y}_2}}$$

$$t = \frac{\text{difference between averages}}{\text{standard deviation of difference}}$$

Case Study: Stat-Ease Bowling Contest



Run	Pat	Mark
1	160	165
2	150	180
3	140	170
4	167	185
5	157	195
6	148	175
Avg	153.7	178.3

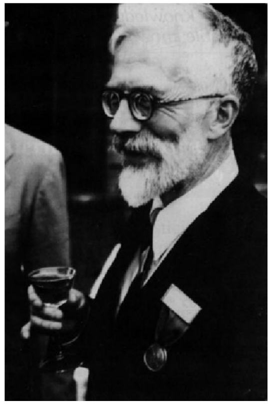
Can say with more than 99.9 % confidence that Mark is a better bowler than Pat based on t = –4.56 (two-tailed p = 0.0004)

Data file: Outtake from "Bowling" tutorial

Multiple Comparisons

Analysis of Variance (ANOVA), the F Test and Least Significant Difference (LSD)

“Personally, the writer prefers to set a low standard of significance at the 5 per cent point, and ignore entirely all results which fail to reach this level. A scientific fact should be regarded as experimentally established only if a properly designed experiment rarely fails to give this level of significance. ”



\- Sir Ronald Fisher

“The Arrangement of Field Experiments,” *The Journal of the Ministry of Agriculture*, 1926, 33, 504.

Little known fact:

When Fisher invented DOE at Rothamsted Experimental Station in England, computations were done by ‘calculators’ – mathematical adepts, mainly female.

Example of Fisher's pioneering work:

A randomized, replicated, blocked experiment *(1/2)*

In a landmark field trial on barley (good for making beer!) in Minnesota, agronomists grew 5 varieties at 5 agricultural stations in 1931 and again in 1932. The study was analyzed by Fisher. *Which variety stands out year-by-year at all locations?*

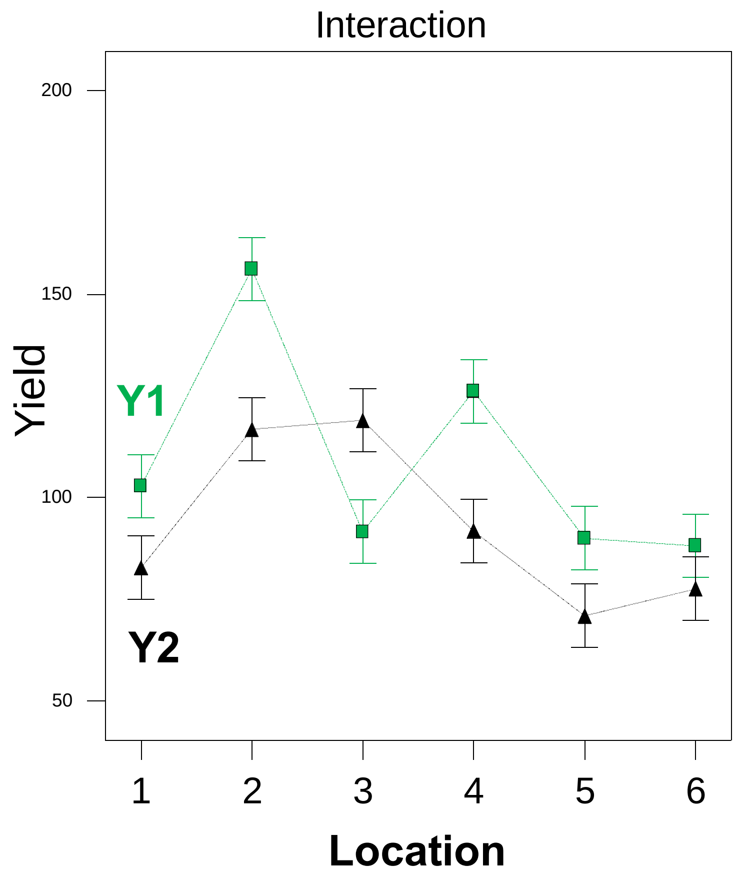
		Varieties				
Location	**Year**	**M**	**S**	**V**	**T**	**P**
1	1	81	105	120	110	98
	2	81	82	80	87	84
2	1	147	145	151	192	146
	2	100	116	112	148	108
3	1	82	77	78	131	90
	2	103	105	117	140	130
4	1	120	121	124	141	125
	2	99	62	96	126	76
5	1	99	89	69	89	104
	2	66	50	97	62	80
6	1	87	77	79	102	96
	2	68	67	67	92	94

Example of Fisher's pioneering work:

A randomized, replicated, blocked experiment *(2/2)*

In a book called *Visualizing Data* (Hobart Press, 1993) William S. Cleveland suggests that the experimenters* reversed the numbers year-by-year in their report for location 3 (Morris, MN). It is hard to see in the raw data, but obvious when graphed with varieties averaged. The 'take home' message:

One picture = 1000 numbers!



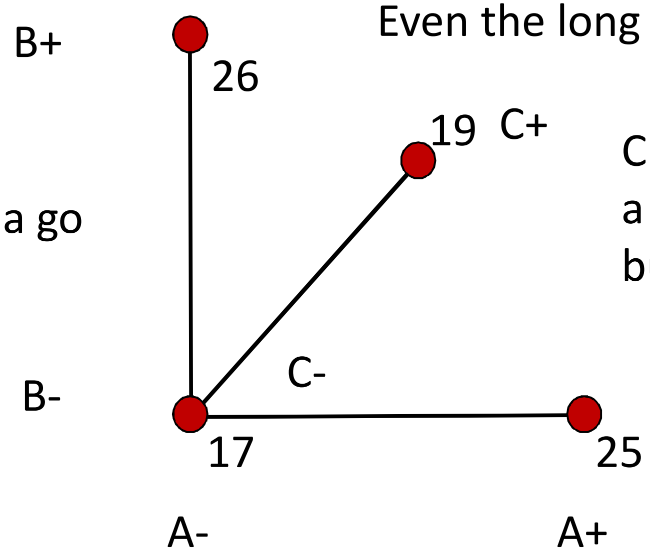
*(Immer, et al, *Journal of Agronomy*, 26, 403-419, 1934).

Data file: Barley

Accelerated Life Test *(1/2)*

No meaningful improvements found with OFAT.

Even the long shot Team C tries



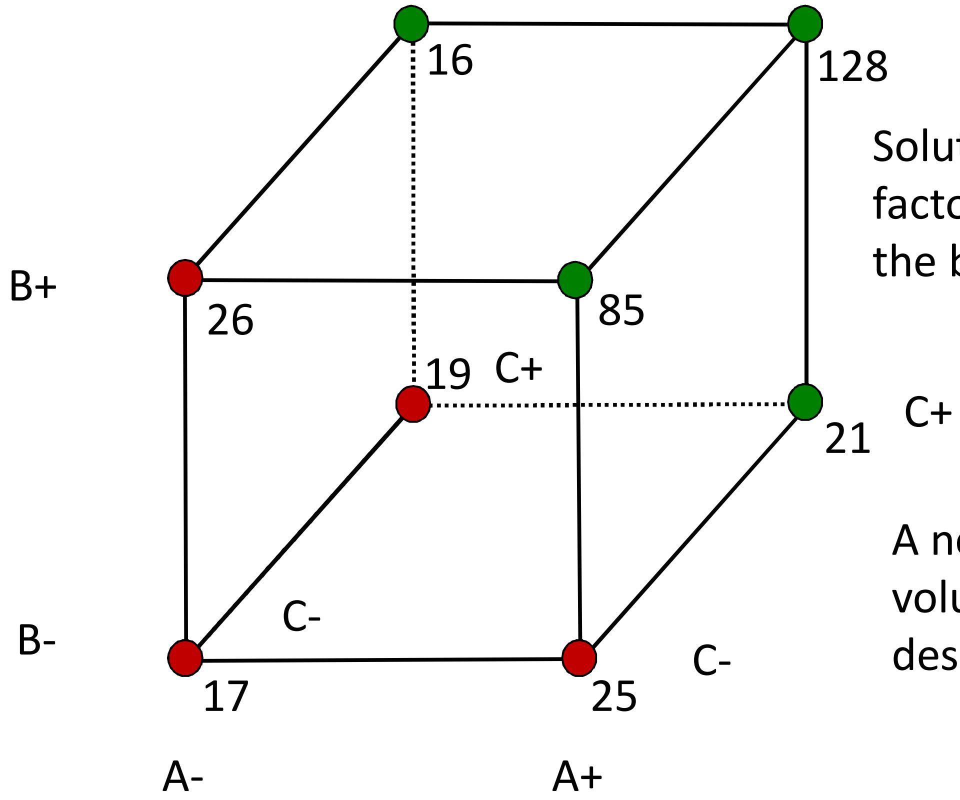
Team B gives it a go

Current operating conditions produce a response of 17 units. But, to stay in business, the response must double.

Team A works on their factor but cannot double the response

Multifactor DOE vs One-Factor at a Time (OFAT)

Accelerated Life Test *(1/2)*

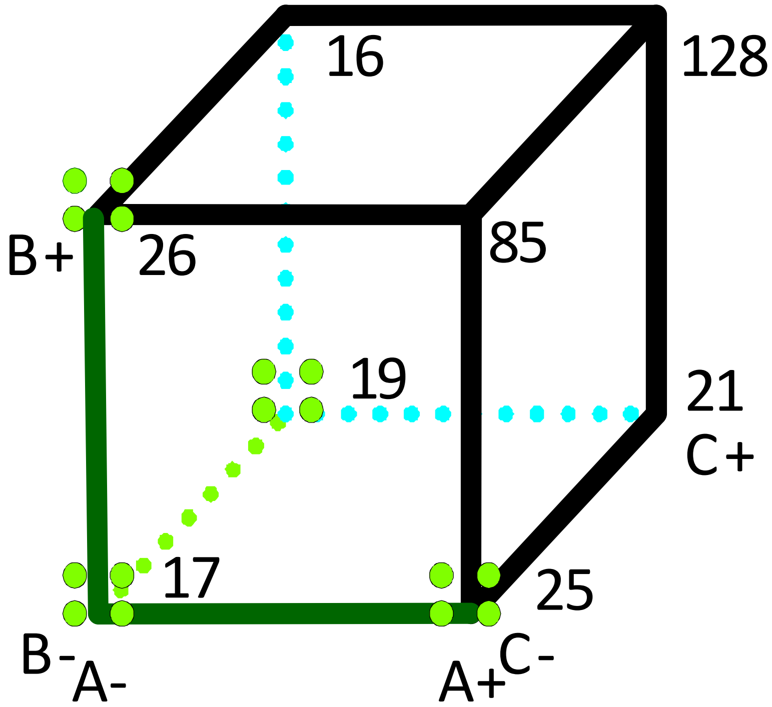


Solution found! Two factors interact to create the breakthrough results.

A newly hired engineer volunteers to do a designed experiment.

Multifactor DOE vs OFAT *(2/2)*

Relative Efficiency for Statistical Power



Relative efficiency = 16/8

□ 2 to 1!

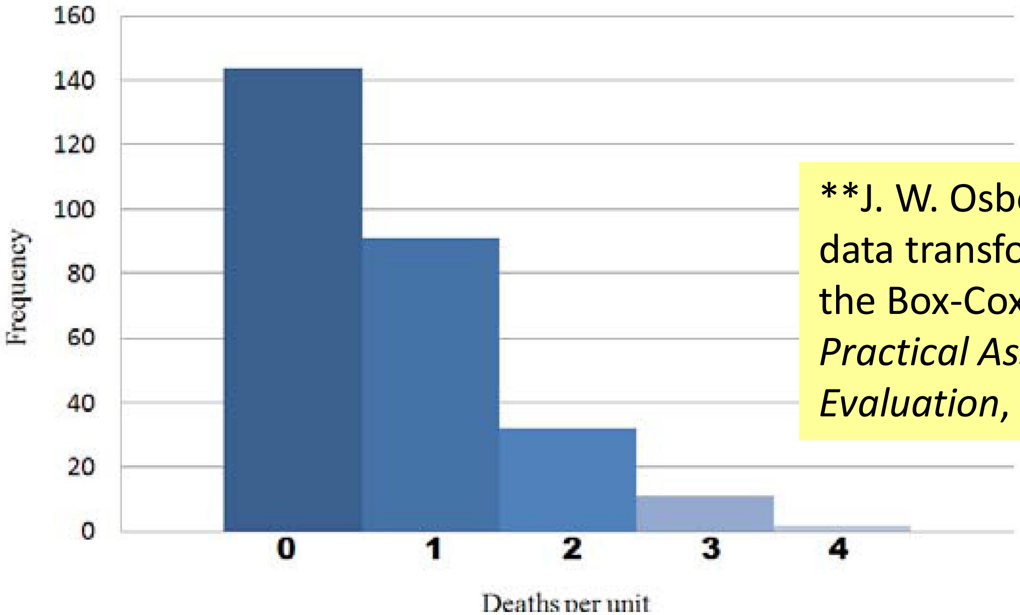
"To make knowledge work productive will be the great management task of this century."

-- Peter Drucker

Box-Cox Transformation*

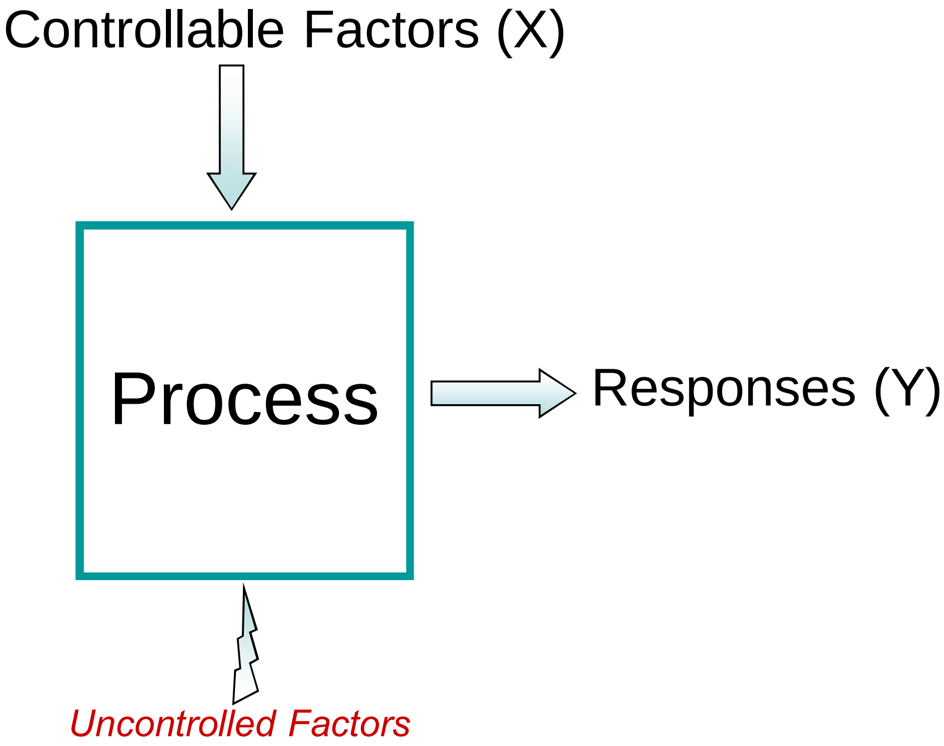
*G. E. P. Box and K. B. Wilson, 1964, "An Analysis of Transformations," *Journal of the Royal Statistical Society*, B26, pp 211-234.

It turns out that the bearing data models far better when transformed by log, which is made perfectly obvious by the Box-Cox plot—a huge innovation. Here's another classic set of data amenable to a Box-Cox transformation—Bortkowiecz's 1898 compilation** on Prussian cavalrymen killed by horse-kicks.



**J. W. Osborne, Improving your data transformations: Applying the Box-Cox Transformation, *Practical Assessment Research & Evaluation*, V15, N12, Oct 2010.

DOE Works on Any Process

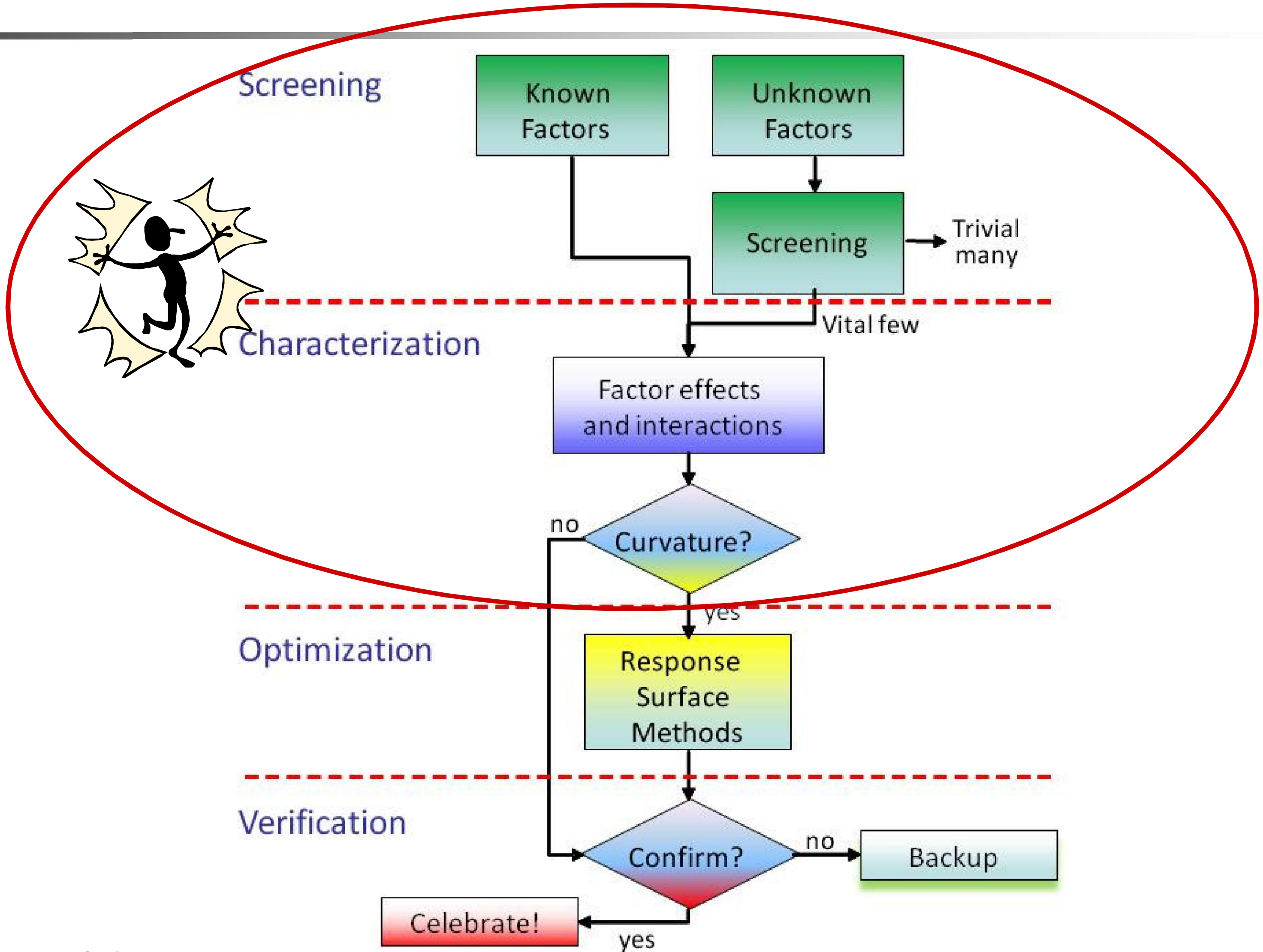


Definition: DOE is:

"A series of tests, in which purposeful changes are made to input factors,

to identify causes for significant changes in the output responses."

Strategy of Experimentation



Screening/Characterization

Purpose: Quickly sift through a large number of potential factors to discard the trivial many. Then follow up with an experiment that focuses on the vital few and how they interact.

Tool: Two-level factorial designs:

1. Fractional for resolving main effects in minimal runs.
2. Full (or less fractional) to resolve two-factor interactions.

Stat-Ease statistics made easy

Two-level factorial case study:

Weed-Whacker Engine

Factor	Name	Low (–)	High (+)
A	Prime pumps	3	5
B	Pulls at choke	3	5
C	Gas at choke	0	100%
D	Final choke	0	50%
E	Gas for start	0	100%

Primer bulb

Choke control

Weed-Whacker Engine:

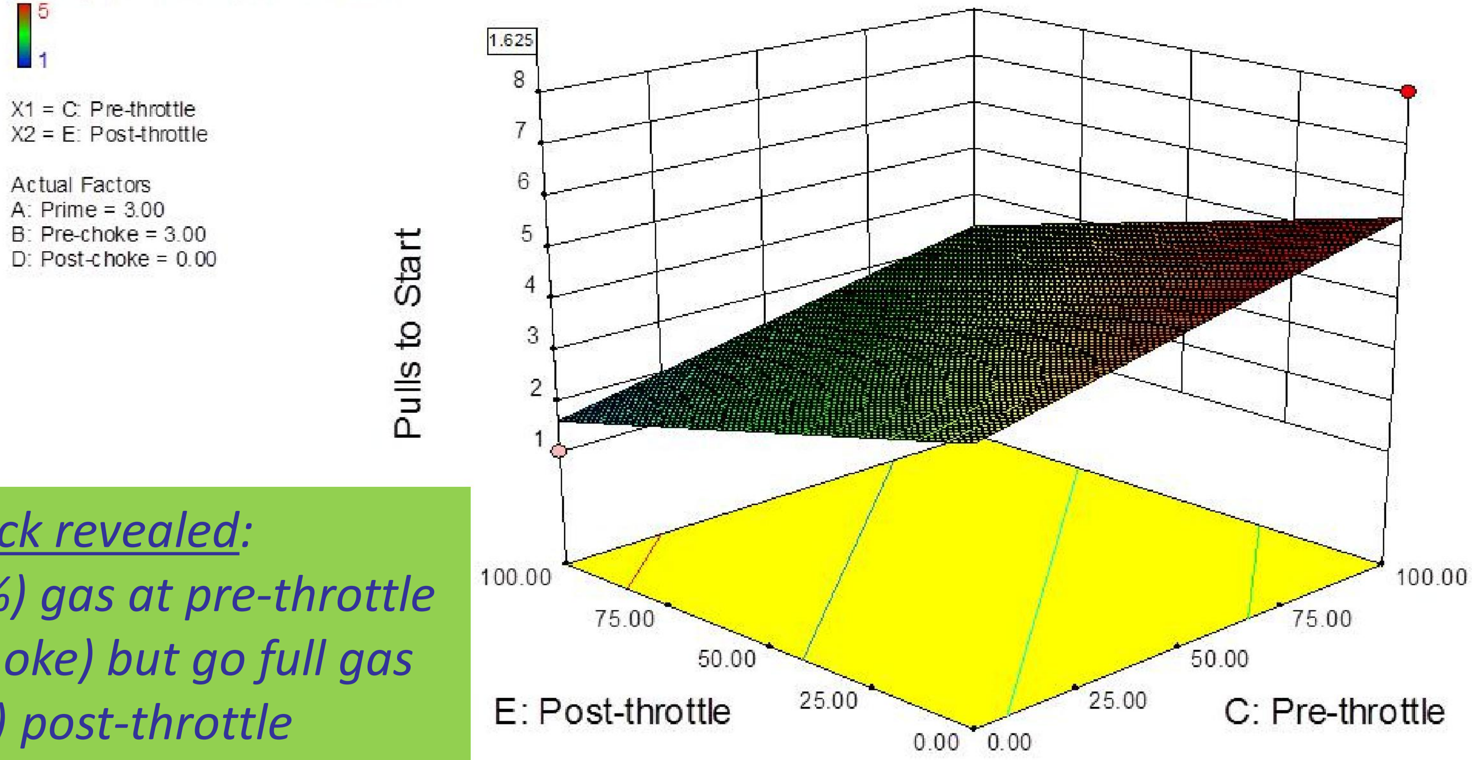
Fractional Factorial Design

Std	A	B	C	D	E	Pulls
1	–	–	–	–	+	1
2	+	–	–	–	–	4
3	–	+	–	–	–	4
4	+	+	–	–	+	2
5	–	–	+	–	–	8
6	+	–	+	–	+	2
7	–	+	+	–	+	3
8	+	+	+	–	–	5
9	–	–	–	+	–	3
10	+	–	–	+	+	1
11	–	+	–	+	+	3
12	+	+	–	+	–	4
13	–	–	+	+	+	3
14	+	–	+	+	–	4
15	–	+	+	+	–	6
16	+	+	+	+	+	5

This factor deliberately aliased to cut runs by half! **[E]=E+ABCD**

This is a standard half-fraction. To achieve equivalent power (8 at high level versus 8 at low), an OFAT experiment on 5 factors requires 3 times as many runs: 48 vs 16, counting 8 at the base point – all low level – against 8 each at high levels for all 5 factors (5x8 = 40).

Ladies and gentleman: Start your engines!



The trick revealed:
No (0%) gas at pre-throttle (full choke) but go full gas (100%) post-throttle (starting engine).

Data File: Weedwhack

Minimum-Run Resolution IV (MR4) Designs*

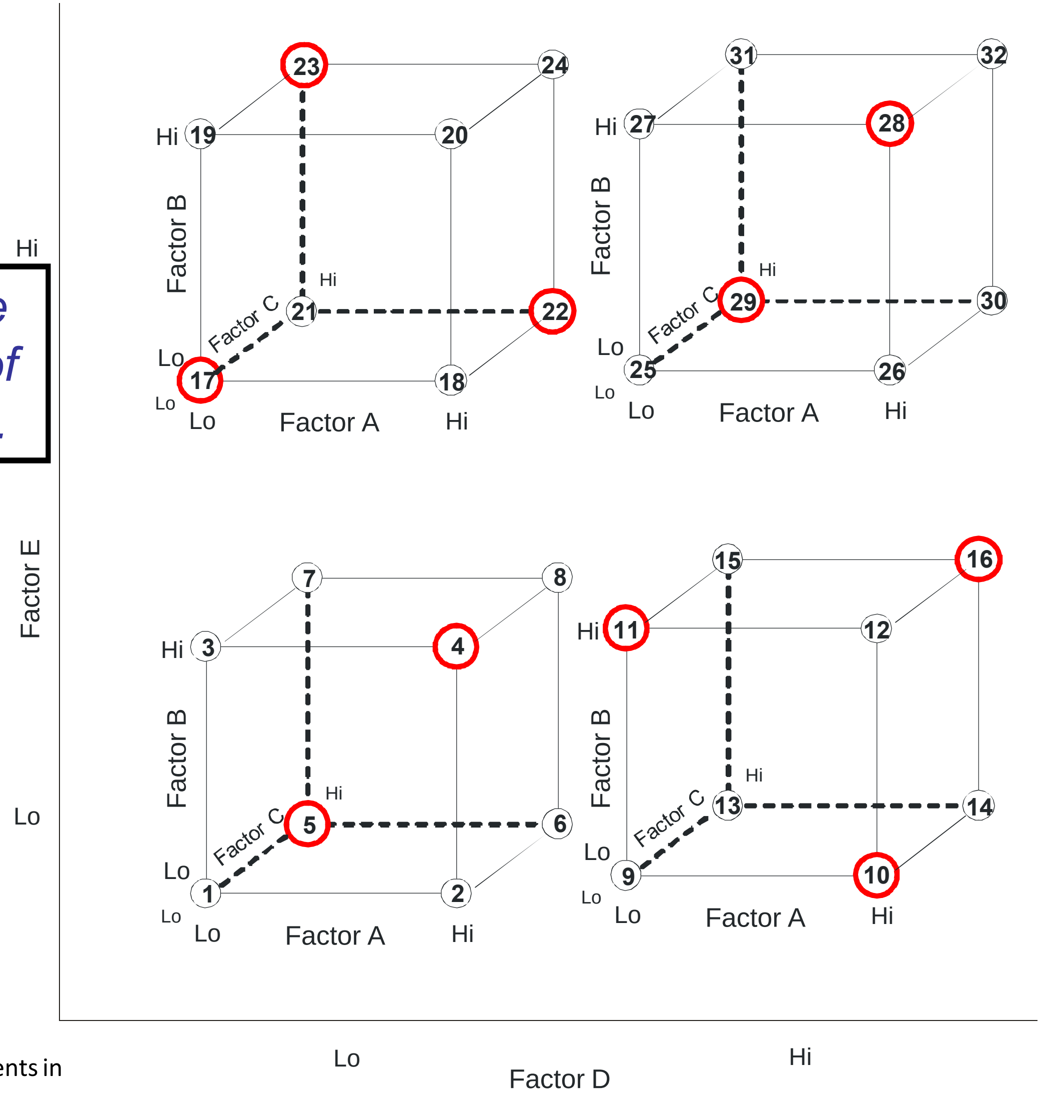
The minimum number of runs for resolution IV design is only two times the number of factors (runs = 2k). This can offer quite a savings when compared to a standard (2^{k-p}) fraction.

> *For example, 9 factors require 32 runs from the standard catalog of designs, but the MR4 requires only 18 (=2x9)!*

*(Anderson & Whitcomb, "Screening Process Factors In the Presence of Interactions," *Annual Quality Congress*, American Society of Quality, Toronto, May, 2004.)

Five-Factor Geometry (10 runs)

Note the balance for every factor of low vs high level.



Minimum-Run Resolution V (MR5) Designs*

NEW!

Regular fractions (2^{k-p} fractional factorials) of 2^k designs often contain considerably more runs than necessary to estimate the coefficients in the 2FI model.

> The smallest regular resolution V design for k=7 uses 64 runs (2^{7-1}) to estimate 29 coefficients.
>
> Our balanced minimum-run resolution V (MR5) design for k=7 has 30 runs, a savings of 34 runs.

* *Small, Efficient, Equireplicated Resolution V Fractions of 2^k designs and their Application to Central Composite Designs*, Gary Oehlert and Pat Whitcomb, 46th Annual Fall Technical Conference, Friday, October 18, 2002.

Minimum-Run Designs (up to 50 factors)

Considerable Savings Over Standard Fractions

NEW!

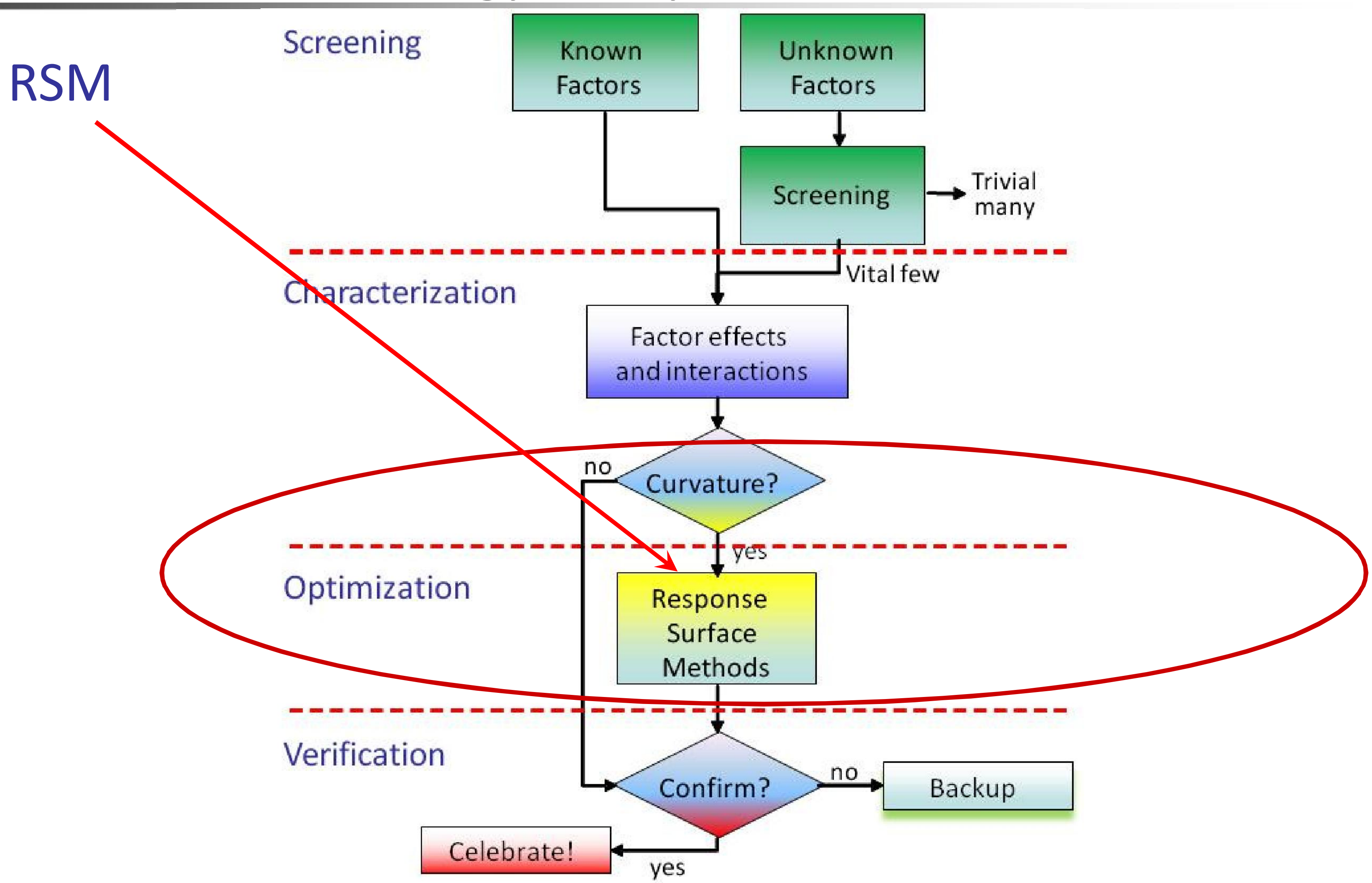
Characterization

Factors	Std Res V	MR5*
6	32	22
7	64	30
8	64	38
9	128	46
10	128	56
11	128	68
12	256	80
13	256	92
14	256	106

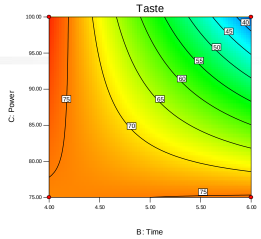
Screening

Factors	Std Res IV	MR4**
9	32	18
10	32	20
11	32	22
12	32	24
13	32	26
14	32	28
15	32	24
16	32	26
17	64	28

Strategy of Experimentation

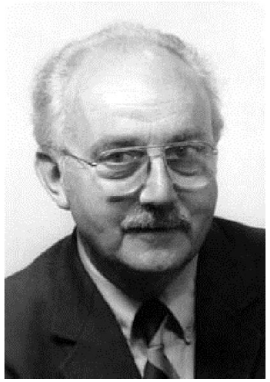


Response Surface Methods (RSM)



Introduced by George E. P. Box and K. B. Wilson in 1951.

(“On the experimental attainment of optimal conditions,”
Journal of the Royal Statistical Society, B13, pp 1-45)



“All models are wrong, but some are useful"

*For a wonderful retrospective see “A Conversation with George Box” by Morris H. DeGroot, Statistical Science, V2, #3 (1987), 239-258. Link to the pdf via http://goo.gl/gW3q6. Also read the 2010 reminiscence by Box about him becoming “An Accidental Statistician” posted by University of Wisconsin at http://www.stat.wisc.edu/~yandell/stat/50-year/Box_George.html.

Response Surface Methods (RSM)*

When to Apply It (Strategy of Experimentation)

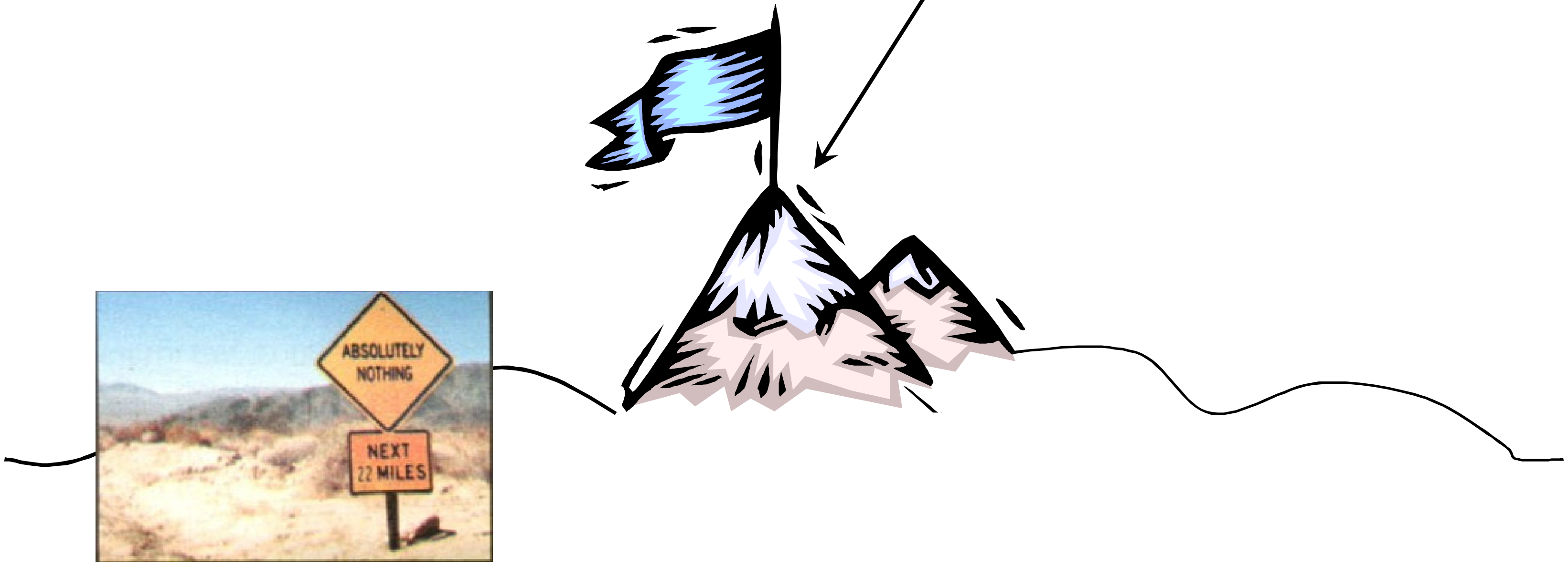
1. Fractional factorials for screening
2. High-resolution fractional or full factorial to understand interactions *(add center points at this stage to test for curvature)*
3. Response surface methods (RSM) to optimize.

Contour maps (2D) and 3D surfaces guide you to the peak.

RSM: When to Apply It

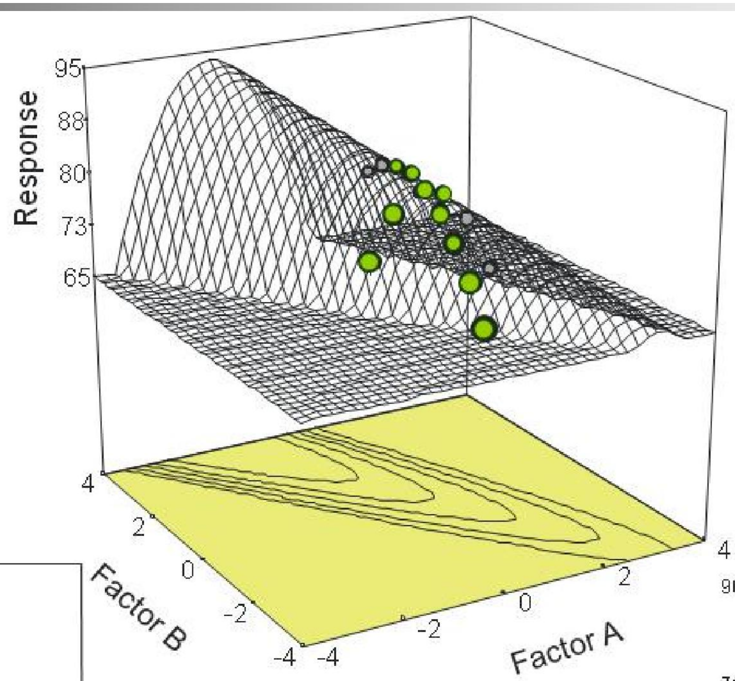
Use factorial design to get close to the peak. Then RSM to climb it.

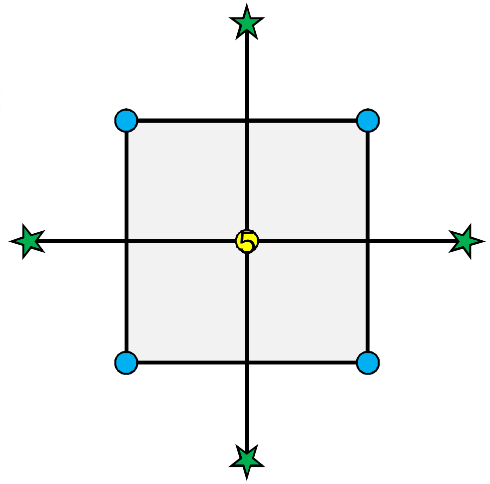
Region of Interest

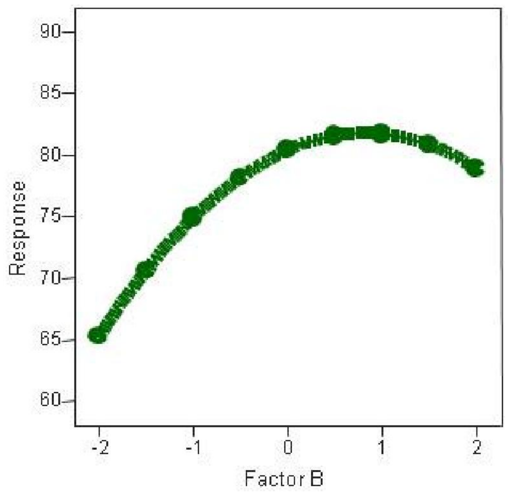


Region of Operability

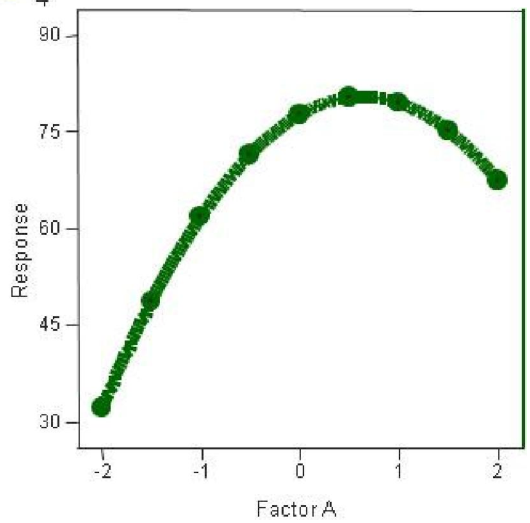
RSM vs OFAT







"People used to say, men may come and men may go, but the yield of this particular product is always 40%." – George Box



RSM: Process Flowchart

Subject Matter Knowledge
(Plus Factorial Screening)

↓

Vital Few Factors (x's)

↓

Process → Measured Response(s) (y(s))

↓

Fitting*

↓

Polynomial Model

↓

Response Surface

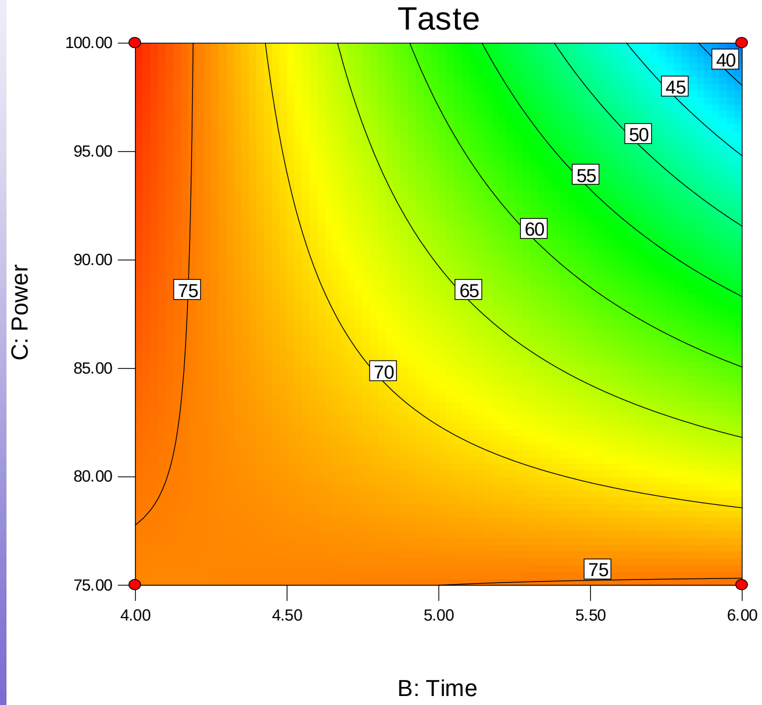
"All models are wrong, but some are useful." - George Box

Simple Example of Design Space

Making Microwave Popcorn *(1/2)*

Where is the design space -- the operating window that provides much taste but few Un-Popped Kernels?

This is a classic trade-off of quality versus yield.



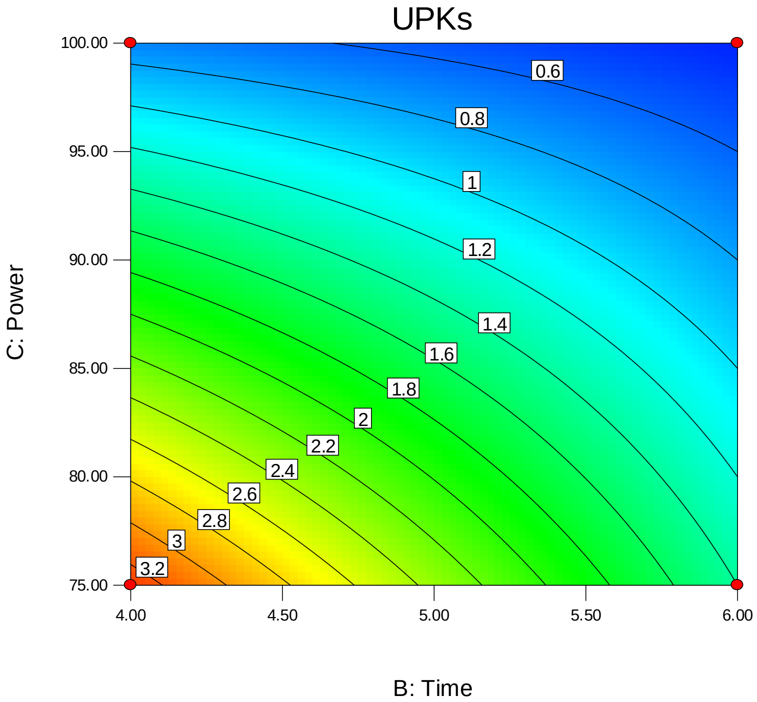
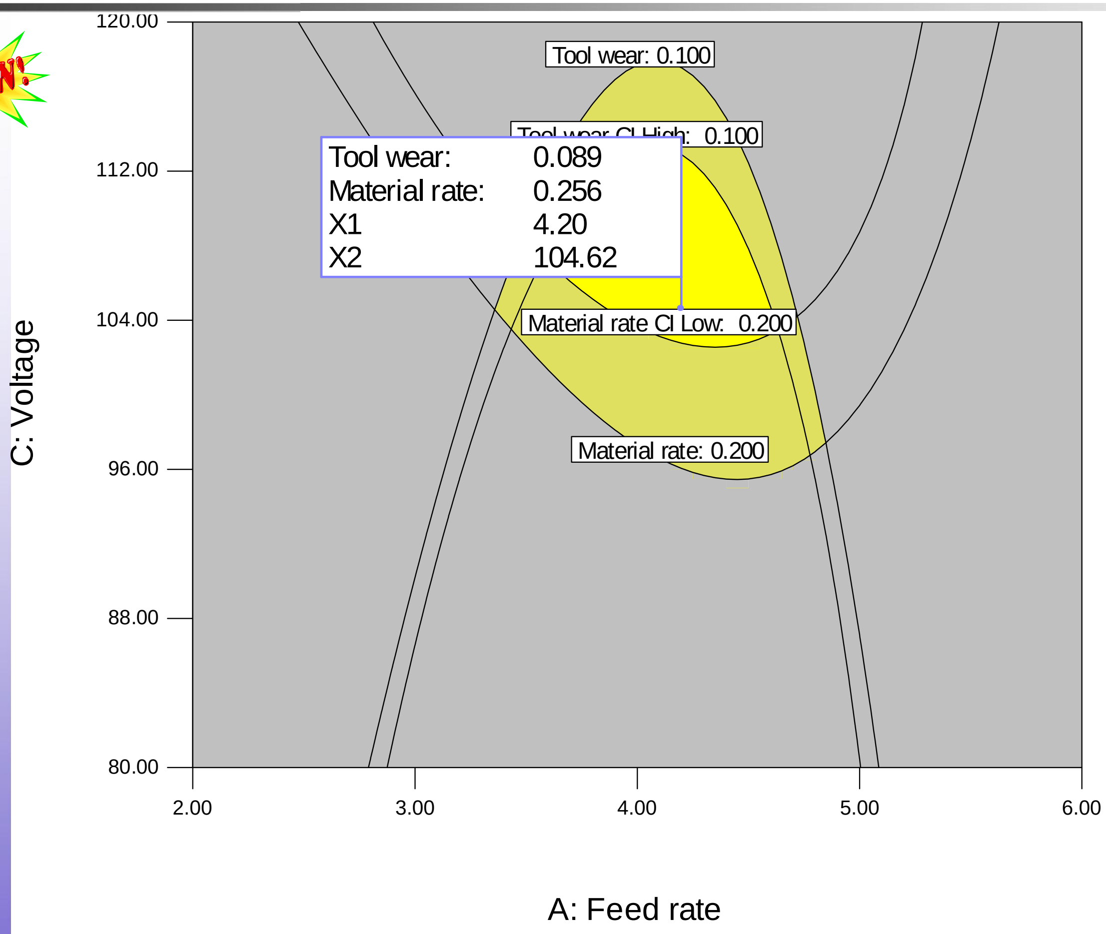


Photo from http://www.ehow.com/info_8028979_science-projects-popcorn-pops.html

Data file: Popcorn

Example of a Functional Design Space

Electrodischarge milling (EDM)



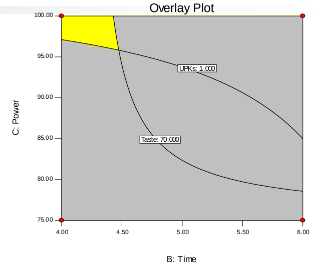
To be conservative (robust) in framing the sweet spot, super-impose the Confidence Intervals (function of under-lying variation and the power of the experiment design). The flag in the center might mark a good place to operate!

"Numerical optimization via desirability investigation of machining parameters for the multiple-response optimization of micro electrodischarge milling," Mehfuz & Ali, *International Journal of Advanced Manufacturing Technology* (2009) 43:264–275

Data file: Milling

Next Big Thing in the Field of DOE/RSM:
Quality by Design (QbD) Design Space

NEW!



By overlaying contour plots for multiple responses – shading out regions out of spec, one can view the <u>design space</u> (aka "operating window" or "sweet spot"), a *"multidimensional combination and interaction of material attributes and process parameters that have demonstrated to provide assurance of quality."* * This is a key element in "quality by design" (QbD). <u>It merits attention from all engineers</u>: chemical, electrical, etc.

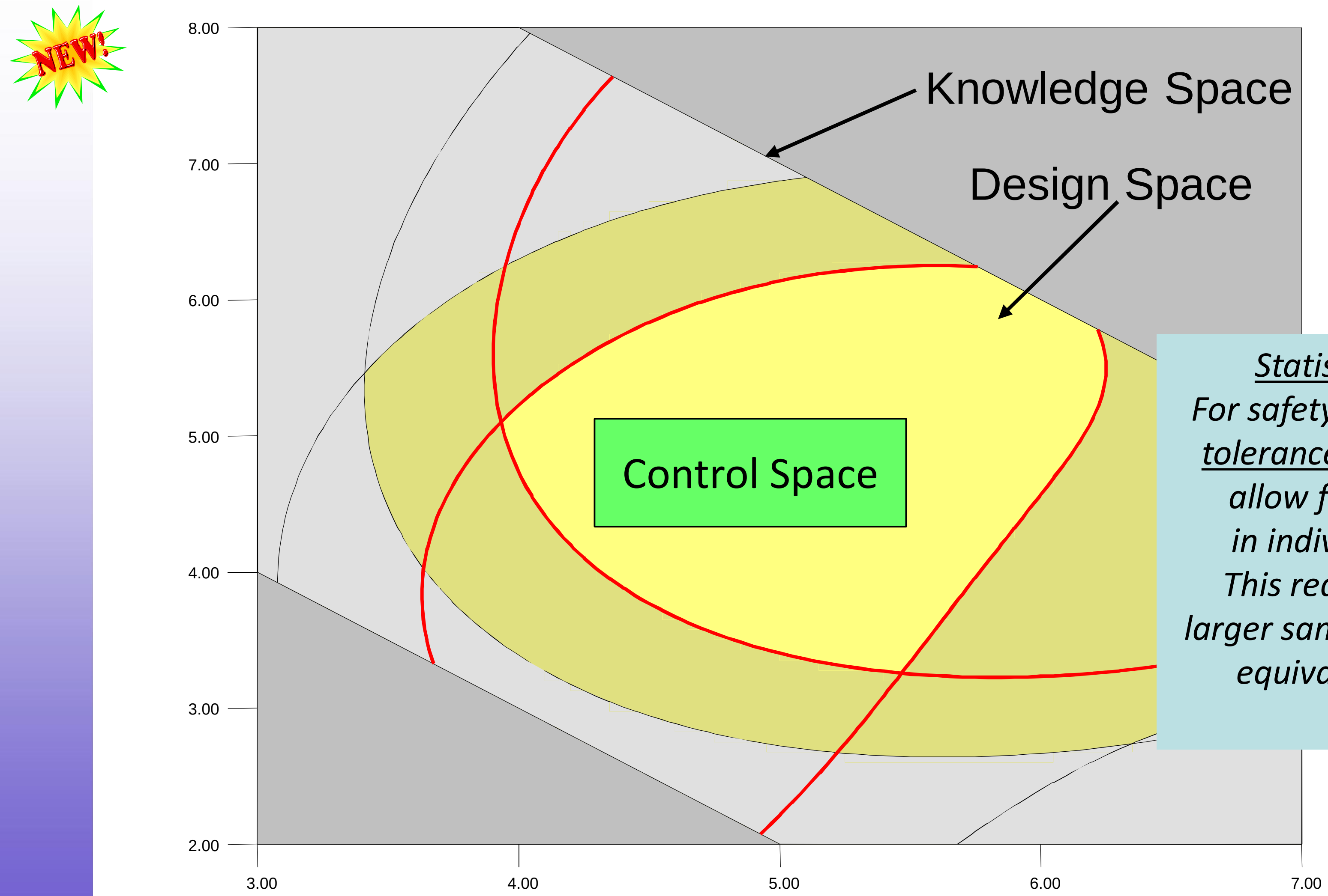
"QbD is a systematic & scientific approach to improve quality, efficiency and profitability."

- Iris Ziegler, Development Director, Nycomed

*(U.S. FDA)

Quality by Design "QbD" Design Space

Allowing for Drift in Your Control Space



<u>Statistical detail</u>:
For safety sake apply <u>tolerance</u> interval to allow for variation in individual units. This requires much larger sample size for equivalent power.

Other Landmark Events in DOE/RSM

Optimal design, Gustav Elving (pictured), 1950
& Box and Lucas 1959 *(continuing to develop)*
("Design of Experiments in Nonlinear Situations," *Biometrika*, 46, pp 77-90.)

Mixture design, Henry Scheffe, 1960s
("The Simplex-Centroid Design for Experiments with Mixtures," *Journal of the Royal Statistical Society* Vol. 25, No. 2, 1963.)

Restricted Maximum Likelihood (REML), 1970s *(for split plots, etc)*
(R. R. Corbeil & S. R. Searle, "Restricted Maximum Likelihood Estimation of Variance Components in the Mixed Model," *Technometrics,*V18, N1, Feb. 1976, pp. 31-38.)

Robust parameter design, Genichi Taguchi, early 1980's

Space filling designs for design and analysis of computer experiments (DACE), late 1980s *(continuing)*
(Sacks, J., Welch, W. J., Mitchell, T. J., and Wynn, H. P. (1989). "Design and Analysis of Computer Experiments," *Statistical Science* 4(4), pp. 409–423.)

Bayesian methods, Box and Meyer, 1992 *(continuing)*
("Finding the Active Factors in Fractionated Screening Experiments," *Technical Report* 80, Center for Quality and Productivity Improvement, University of Wisconsin.)

Conclusion (after 75 years!)

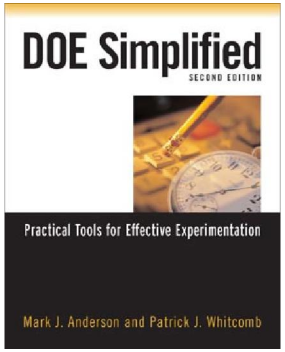
Trim out the OFAT!

By making use of <u>multifactor</u> design of experiments (DOE) starting with simple two-level factorials and graduating to response surface methods (RSM) for processes, you will greatly accelerate product development and process optimization.

Postcript: Things on the Horizon *(1/2)*

VODcasts (video on demand) and other forms of digital delivery are revolutionizing how statistics get taught*

*“Students attend lectures virtually on the Internet”, p. 13,
The Global Edition of the New York Times, Monday, June 25, 2012



Web-Based
‘Launch Pad’
(Chapters 1-3)
http://statease.info
/DOEpresentations/
(view 1-2 with PIP

”In 50 years, there will be only 10 institutions in the world delivering higher education.”

- *Wired Magazine,* March 20, 2012, “The Stanford Education Experiment Could Change Higher Learning Forever”, Steven Leckart
http://www.wired.com/wiredscience/2012/03/ff_aiclass/all/

Postcript: Things on the Horizon *(2/2)*

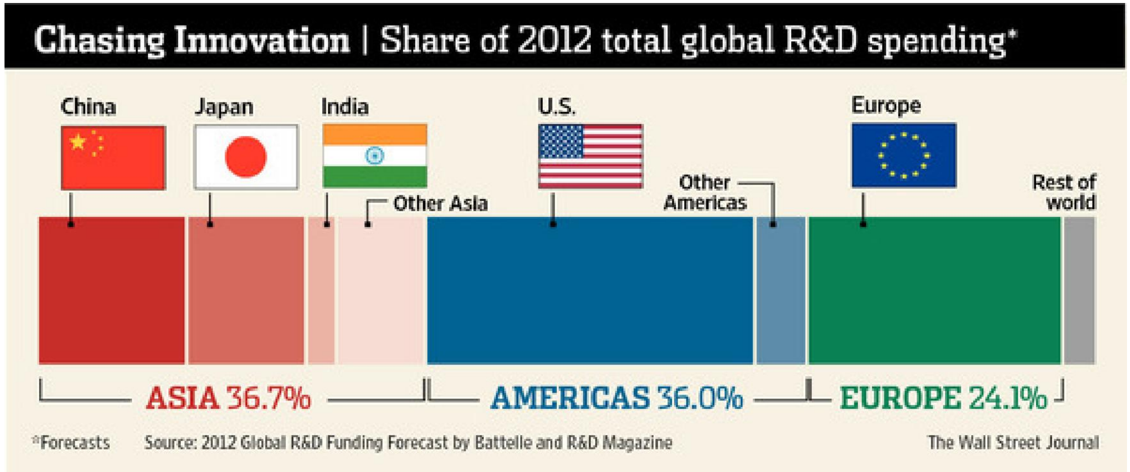
Statistical Engineering

"The application of statistical inference to engineering experiments" per Merriam-Webster, but much more to it than that-- see

- http://asq.org/statistics/quality-information/statistical-engineering
- Statistical Engineering Special Issue, *Quality Engineering*, 24, 2012.

What tools will be used to win the worldwide R&D battle?

Discuss at break and remainder of our meeting!



"There's no statistically significant relationship between how much a company spends on R&D and how they perform over time. What matters most is the kind of innovator you are."
Wall Street Journal, "Myths of the Big R&D Budget" 6/15/12

Best of luck for your multifactor experimenting!

Thanks for listening!

-- Mark

Mark J. Anderson, PE, CQE
Stat-Ease, Inc.
mark@statease.com