MIAMI UNIVERSITY

VCU

A Meta Analysis of Response Surface Experiments

Presented by: **Byran Smucker**

Rebecca Ockuly

Maria Weese

David Edwards

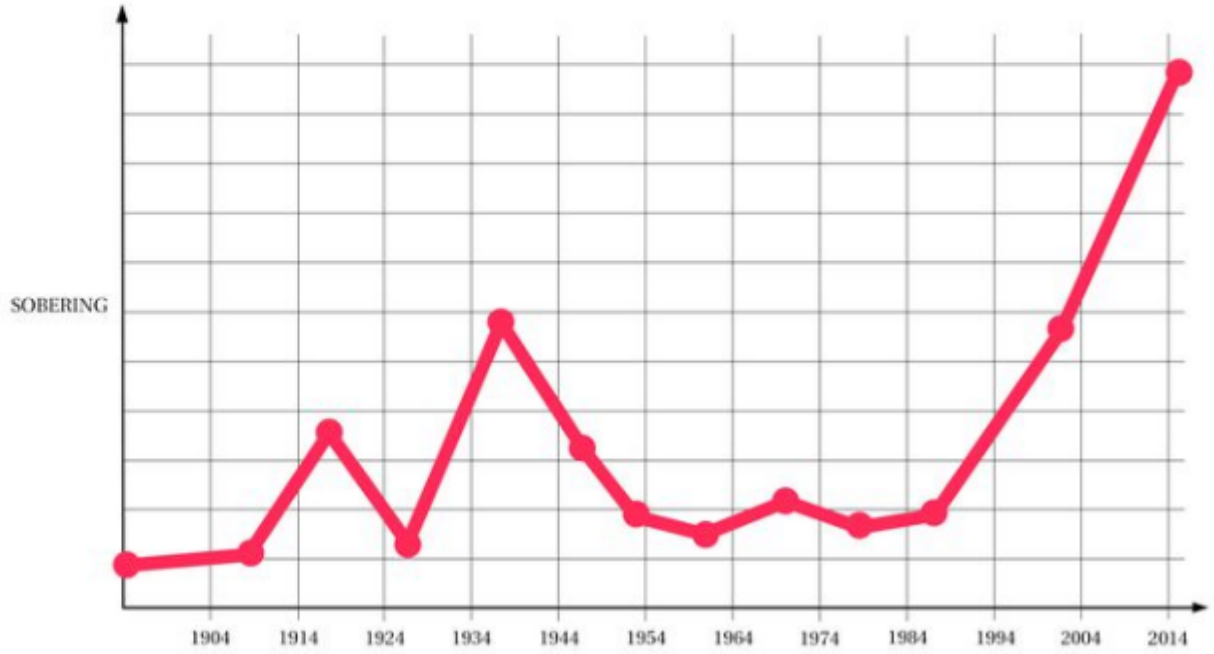
Chang Le

College of Arts & Science

Department of Statistics

Study Finds 79% Of Statistics Now Sobering

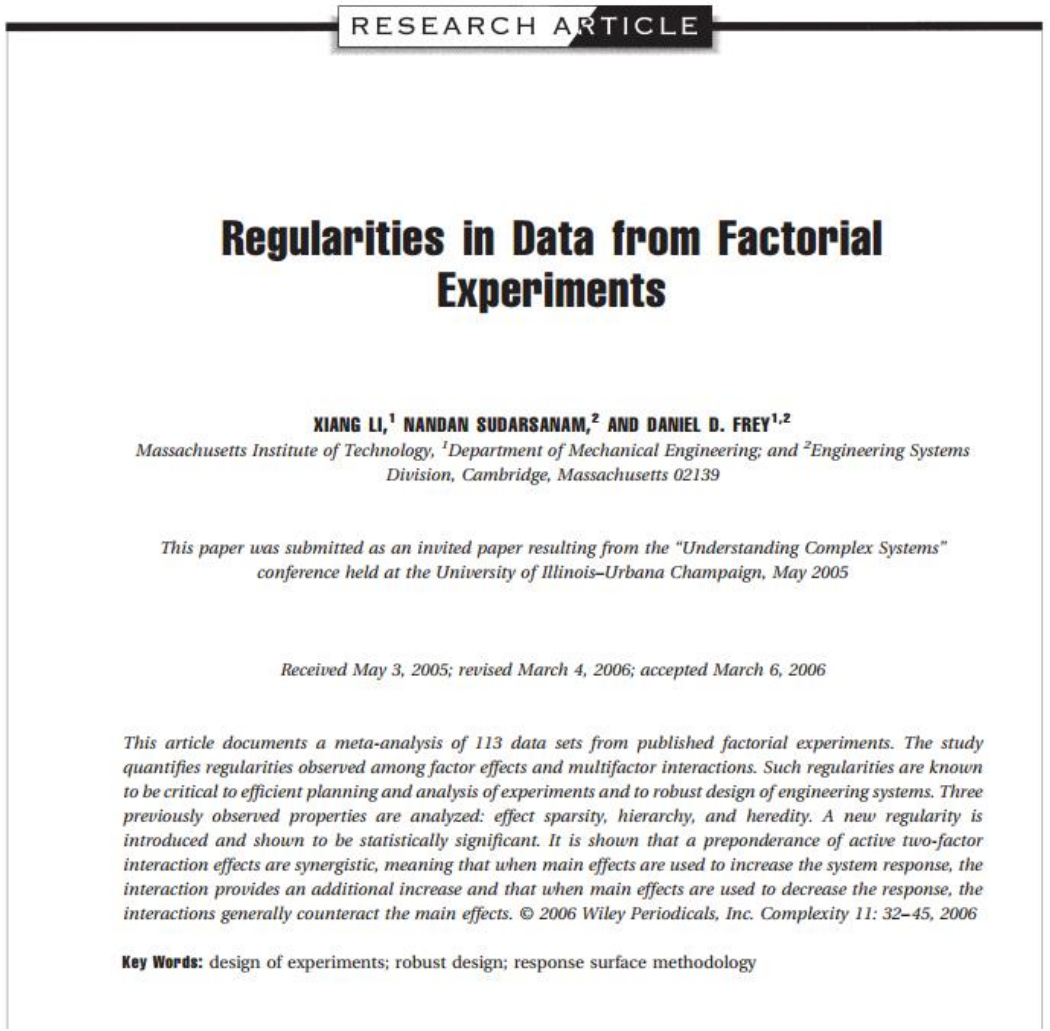
NEWS IN BRIEF
August 20, 2014
VOL 50 ISSUE 33
News



My own satirical headline: Professor scared of being scooped on research project no one else cares about

Motivation

Li, Sudarsanam and Frey (2006) empirically study effect **sparsity**, **heredity**, and **hierarchy** for factorial designs.



RESEARCH ARTICLE

Regularities in Data from Factorial Experiments

XIANG LI,[1] NANDAN SUDARSANAM,[2] AND DANIEL D. FREY[1,2]
Massachusetts Institute of Technology, [1]Department of Mechanical Engineering; and [2]Engineering Systems Division, Cambridge, Massachusetts 02139

This paper was submitted as an invited paper resulting from the "Understanding Complex Systems" conference held at the University of Illinois–Urbana Champaign, May 2005

Received May 3, 2005; revised March 4, 2006; accepted March 6, 2006

This article documents a meta-analysis of 113 data sets from published factorial experiments. The study quantifies regularities observed among factor effects and multifactor interactions. Such regularities are known to be critical to efficient planning and analysis of experiments and to robust design of engineering systems. Three previously observed properties are analyzed: effect sparsity, hierarchy, and heredity. A new regularity is introduced and shown to be statistically significant. It is shown that a preponderance of active two-factor interaction effects are synergistic, meaning that when main effects are used to increase the system response, the interaction provides an additional increase and that when main effects are used to decrease the response, the interactions generally counteract the main effects. © 2006 Wiley Periodicals, Inc. Complexity 11: 32–45, 2006

Key Words: design of experiments; robust design; response surface methodology

Motivation

But what about regularities for response surface experiments, when a second-order model is fit?

$$Y = \beta_0 + \sum_{i=1}^{k} (\beta_i X_i + \beta_{ii} X_i^2) + \sum_{i=1}^{k-1} \sum_{j=i+1}^{k} \beta_{ij} X_i X_j + \epsilon$$

But what about regularities for response surface experiments?

We want to answer questions like:

1. Do the sparsity, heredity, and hierarchy results for factorials hold up for main effects and interactions in RSM?
2. What about applying these principles to quadratic effects?

Objectives

- Refine experimenter expectations
- Provide information to help researchers construct more efficient designs
- Furnish data to inform simulations from RSM models

Terminology

Effect Sparsity: Only a small proportion of terms in a full quadratic model are active.

Effect Hierarchy: Lower-order effects are typically larger than higher-order effects; hence, lower-order effects are more likely to be active than higher-order effects.

Effect Heredity: In order for an interaction to be active, one or both of its parent factors should also be active; similarly, in order for a quadratic term to be active, its parent factor(s) should be active.

Population of RSM Studies Considered

We used the Web of Science API to search all journal articles with the following characteristics:

- Published between January 1st 1990 and December 31st 2014 in the Science Citation Index and the Social Sciences Citation Index.
- Searched on terms “Response Surface” OR “Central Composite” OR “Box Behnken” OR “Box-Behnken”.

This returned 24,286 search results from which we extracted the citations.

The Sample (1/5)

We used a stratified random sample of 129 papers (263 responses) from the population.

Design type: 101 CCDs and 28 BBDs

Runs: 9 to 100 with a mode=32 and a median=29

Axial distance (for CCDs): 0.6 to 2.83, mode=2, median=2 (zero face-centered CCD's!)

Lack of fit (based on responses): 89 show LOF, 166 do not, and LOF could not be estimated in 8.

The Sample (2/5)

We used a stratified random sample by adding “five factors”, “six factors”, and “seven factors” in subsequent searches.

		Papers from WoS	Papers Sampled	Usable Papers	Approx. Popn Size
Sampled Category	Two-Four	24,124	188	67	8,600
	Five	119	89	42	56
	Six	22	22	8	8
	Seven	21	21	12	12
	Total	24,286	320	129	8,676

The Sample (3/5)

The sampled category was not necessarily the actual category.

		Actual Category			
		Two-Four	Five	Six	Seven
Sampled Category	Two-Four	63	3	1	0
	Five	8	34	0	0
	Six	4	0	4	0
	Seven	10	0	0	2
	Total	85	37	5	2

The Sample (4/5)

If papers had multiple responses for a given experiment, we used them.

		Papers	**Responses**	**Responses/Paper**
Actual Category	Two-Four	85	182	2.14
	Five	37	71	1.92
	Six	5	6	1.2
	Seven	2	3	1.5

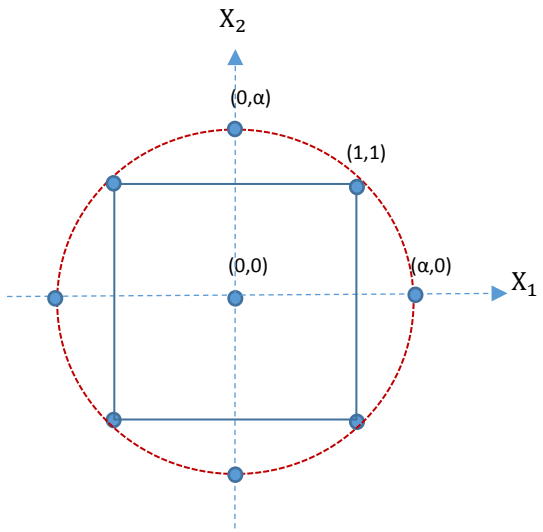
The Sample (5/5)

Weights, in the last column, used to compute population estimates of sparsity, hierarchy, and heredity.

	Weight (sampled)	**Weight (actual)**	**Weight (final)**
Two-Four	0.9912	0.9349	0.9462
Five	0.0065	0.0496	0.0450
Six	0.0009	0.0153	0.0087
Seven	0.0014	0.0002	0.0002

Example (1/2)

CCD with 2 factors and 1 response and n=13 runs.



	Factors		Response
	NaCl (X1)	CaSO (X2)	Lipase Activity (Y)
Factorial Points	-1	-1	2.8
	-1	1	2.6
	1	-1	2.9
	1	1	2.2
Axial Points	-1.414	0	2.1
	1.414	0	2.4
	0	-1.414	2.7
	0	1.414	2
Center Points	0	0	3.3
	0	0	2.9
	0	0	3.2
	0	0	3.5
	0	0	3.3

Liu, Chien-Hung, Wei-Bin Lu, and Jo-Shu Chang. 2006. "Optimizing Lipase Production of *Burkholderia* Sp. by Response Surface Methodology." *Process Biochemistry* 41 (9). Elsevier: 1940–44.

Example (2/2)

CCD with 2 factors and 1 response and n=13 runs.

Define active effect in two ways:

1. Using a *p-value* < 0.05
2. Using a False Discovery Rate (FDR) adjusted *p-value* < 0.05.

We used the absolute value of the t-statistics to quantify effect size.

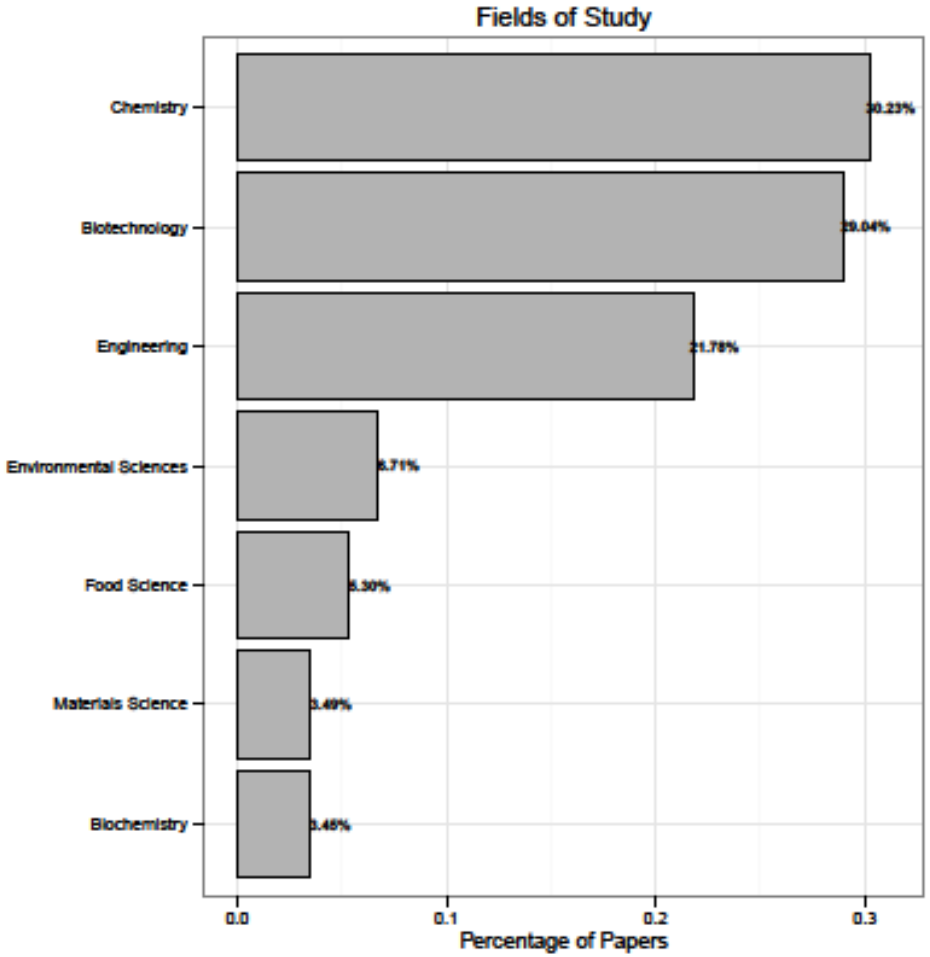
	Estimate	Std. Error	t-value	p-value	FDR p-value
Intercept	3.240	0.116	27.955	1.925e-08	1.155e-07
X1	0.016	0.092	0.170	0.870	0.870
X2	-0.236	0.092	-2.578	0.037	0.056
X1:X2	-0.125	0.130	-0.965	0.367	0.440
X1^2	-0.414	0.098	-4.210	0.004	0.012
X2^2	-0.364	0.098	-3.701	0.008	0.016

	DF	Sum Sq	Mean Sq	F-value	p-value
FO(X1, X2)	2	0.448	0.224	3.338	0.096
TWI(X1, X2)	1	0.063	0.063	0.931	0.367
PQ(X1, X2)	2	1.870	0.935	13.919	0.004
Residuals	7	0.470	0.067		
Lack of Fit	3	0.278	0.093	1.932	0.266
Pure Error	4	0.192	0.048		

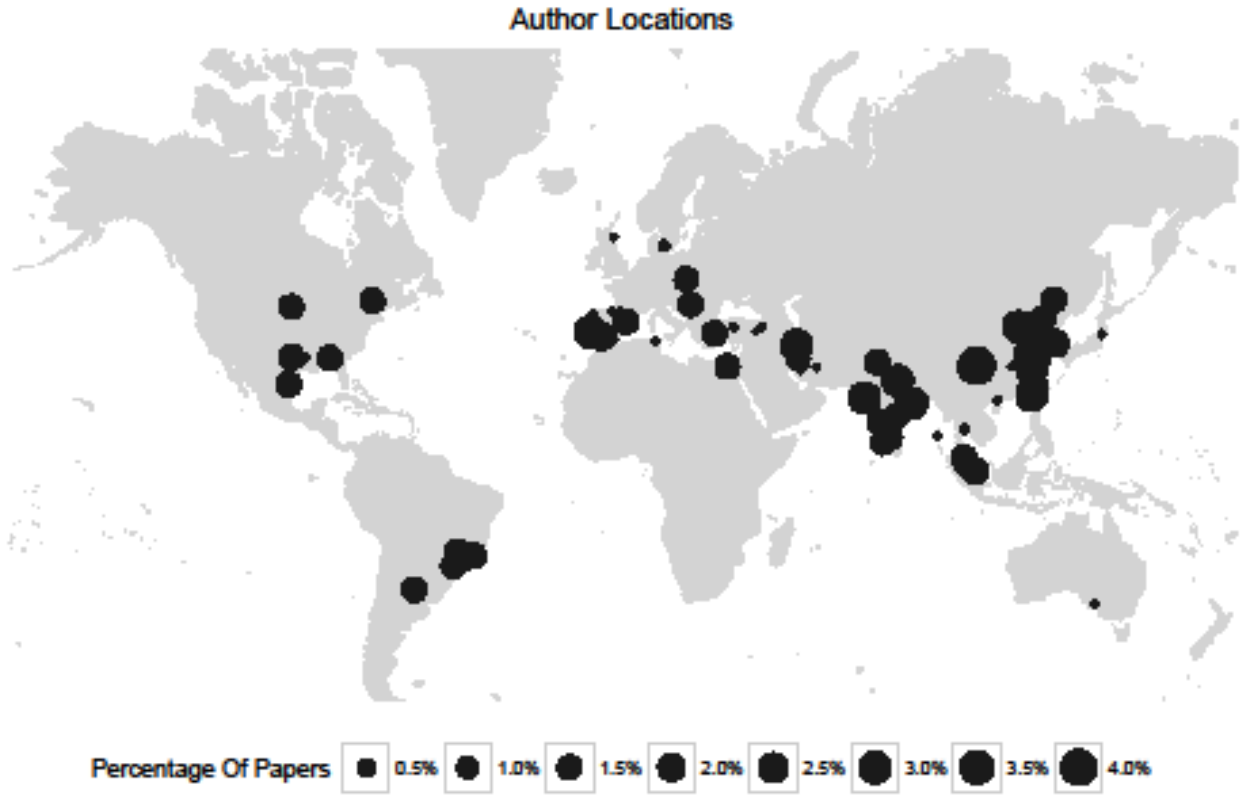
Stationary Point: (0.070, -0.337)

Liu, Chien-Hung, Wei-Bin Lu, and Jo-Shu Chang. 2006. "Optimizing Lipase Production of *Burkholderia* Sp. by Response Surface Methodology." *Process Biochemistry* 41 (9). Elsevier: 1940–44.

EDA (1/2)


Fields of Study
Chemistry
Biotechnology
Engineering
Environmental Sciences
Food Science
Materials Science
Biochemistry
30.23%
29.04%
21.78%
6.71%
5.30%
3.49%
3.45%
0.0
0.1
0.2
0.3
Percentage of Papers

EDA (2/2)



Effect Sparsity (1/2)

p-value < 0.05

Effect Type	Proportion Active	Standard Error
Main Effects	0.58	0.03
Interaction Effects	0.21	0.02
Quadratic Effects	0.45	0.03

FDR p-value < 0.05

Effect Type	Proportion Active	Standard Error
Main Effects	0.51	0.03
Interaction Effects	0.16	0.02
Quadratic Effects	0.39	0.03

Proportions and SE's adjusted to account for stratified sample.

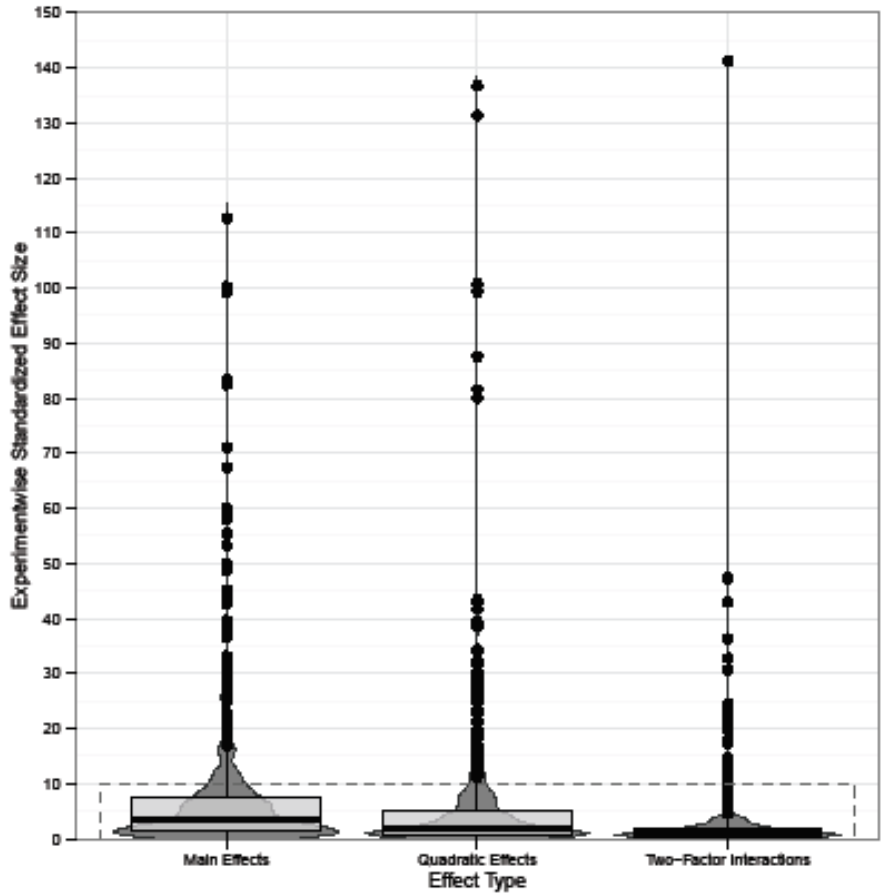
Li, Sudarsanam and Frey (2006) found 41% of main effects to active and 11% of two factor interactions to be active.

Effect Sparsity (2/2)

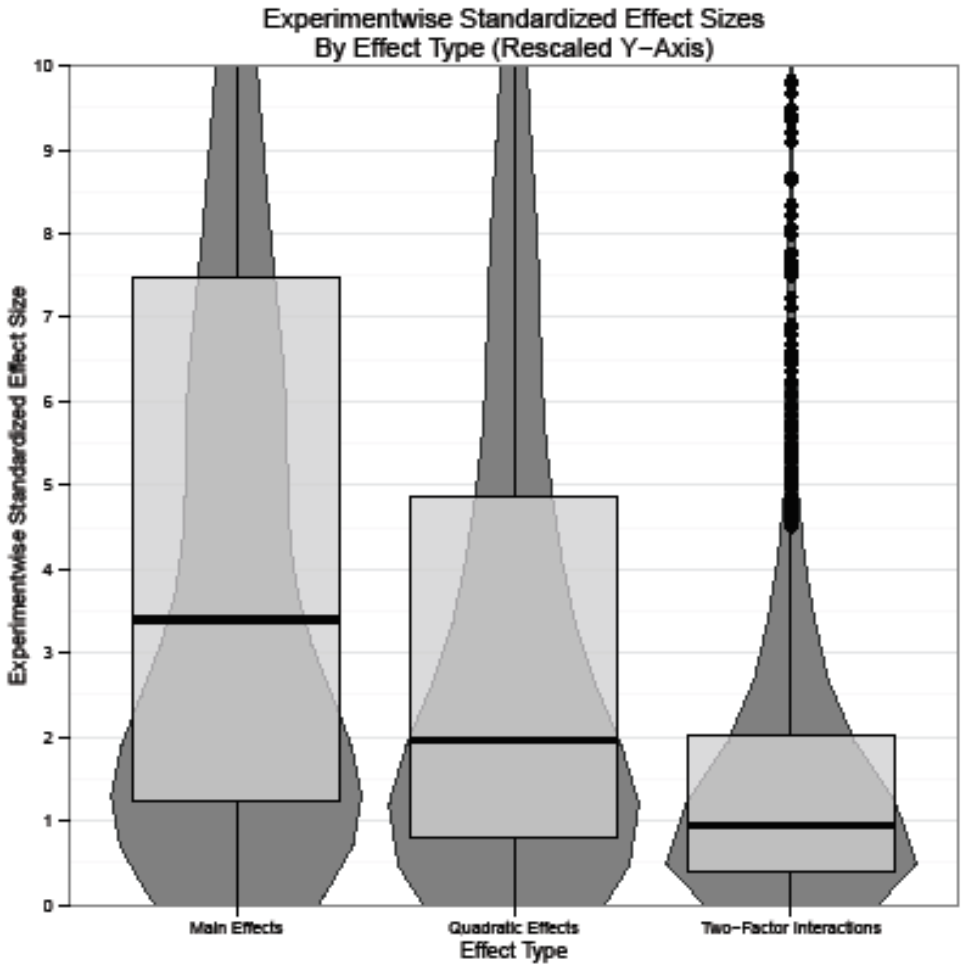
k	Number of Responses	Main Effects (SE)	Two-Factor Interactions (SE)	Quadratic Terms (SE)
2	48	0.54 (0.05)	0.15 (0.05)	0.36 (0.05)
3	94	0.53 (0.03)	0.21 (0.02)	0.42 (0.03)
4	40	0.65 (0.04)	0.21 (0.03)	0.53 (0.04)
5	71	0.54 (0.03)	0.23 (0.02)	0.51 (0.03)
6	6	0.64 (0.08)	0.22 (0.04)	0.74 (0.07)
7	3	0.62 (0.11)	0.29 (0.06)	0.38 (0.11)

Based on unadjusted p-values. Similar story for the FDR-adjusted p-values.

Effect Hierarchy (1/3)



Effect Hierarchy (2/3)



Effect Hierarchy (3/3)

Effect Type	Mean Effect Size (SE)	Median Effect Size (SE)
Main Effects	6.26 (0.45)	3.39 (0.31)
Interaction Effects	1.62 (0.09)	0.94 (0.05)
Quadratic Effects	3.71 (0.20)	1.97 (0.20)

All effects are included, not just active effects.

Li, Sudarsanam and Frey (2006) reported that the median of the main effects to be about 4 times larger than the median of the two factor interactions.

Effect Heredity (1/3) - Traditional

$\hat{P}(\textit{active interaction} | \textit{active main effect(s)})$

p-value < 0.05

Strength	Example	Proportion	Std. Error
Strong	AB, A, B	0.36	0.04
Weak	AB, A	0.15	0.03
None	AB	0.05	0.02

FDR p-value < 0.05

Strength	Example	Proportion	Std. Error
Strong	AB, A, B	0.35	0.04
Weak	AB, A	0.11	0.03
None	AB	0.03	0.01

Proportions adjusted to reflect stratified sampling

Li, Sudarsanam and Frey (2006) reported 33% (strong), 4.5% (weak) and 0.48% (none)

Effect Heredity (2/3) - with Quadratic Effects

$\hat{P}(\textit{active quadratic effect} \mid \textit{active main effect})$

p-value < 0.05

Strength	Example	Proportion	Std. Error
Heredity	A^2, A	0.56	0.04
No Heredity	A^2	0.30	0.04

FDR p-value < 0.05

Strength	Example	Proportion	Std. Error
Heredity	A^2, A	0.54	0.04
No Heredity	A^2	0.23	0.04

$\hat{P}(\textit{active interaction} \mid \textit{active quadratic effect(s)})$

Strength	Example	Proportion	Std. Error
Strong	AB, A^2, B^2	0.31	0.05
Weak	AB, A^2	0.26	0.04
None	AB	0.08	0.02

Strength	Example	Proportion	Std. Error
Strong	AB, A^2, B^2	0.31	0.05
Weak	AB, A^2	0.20	0.04
None	AB	0.06	0.02

Effect Heredity (3/3) - Expanded

$\hat{P}(\textit{active interaction} | \textit{active main effect(s) and quadratic effects})$

Strength	Example	Unadjusted (SE)	FDR-adjusted (SE)
Strong-Strong	AB, A, B, A^2, B^2	0.45 (0.06)	0.43 (0.06)
Strong-Weak	AB, A, B, A^2	0.39 (0.06)	0.37 (0.07)
Strong-None	AB, A, B	0.19 (0.05)	0.20 (0.04)
Weak-Strong	AB, A, A^2, B^2	0.24 (0.06)	0.22 (0.07)
Weak-Weak	AB, A, A^2 OR AB, A, B^2	0.18 (0.04)	0.09 (0.04)
Weak-None	AB, A	0.05 (0.03)	0.04 (0.03)
None-Strong	AB, A^2, B^2	0.05 (0.04)	0.10 (0.06)
None-Weak	AB, A^2	0.08 (0.05)	0.06 (0.05)
None-None	AB	0.05 (0.03)	0.01 (0.01)

Miscellaneous

Interaction types:

- 28% of interactions were *reinforcing* – interaction same sign as both main effects;
- 28% *interfering* – interaction opposite sign as both main effects;
- 44% have one parent positive and one negative.

Stationary points (adjusted for stratification):

- About 50% inside design region
- About 32% maxima, 14% minima, and 55% saddle points

Concerns

- If you look hard enough at an applied RSM paper, you'll find something not to like; e.g. mysterious decisions about screening
- For 6- and 7- factors experiments, we ostensibly sampled entire population; for 5-factors we sampled most of it. This led to some curious results:
  - Three pairs of papers with same design matrix, but different response values
  - One pair of papers that shared a response, with same response values

Satirical headline: Statistician examines every paper in the scientific literature; finds errors in all but those he co-authored.

Caveats

- Treated each response as an independent experiment.
- Limited to the population of published experiments (with data) from Web of Science between 1990 and 2015.

Conclusions

- Effect Sparsity: **Main effects** and **quadratic effects** are more common than **interactions**
- Effect Hierarchy: **Interaction effects** are smallest; **quadratic effects** are about twice the size of interactions; **main effects** are about twice the size of quadratic effects
- Effect Heredity: More complicated to summarize, but strong and weak heredity certainly exist for both **main effects** and **quadratic effects.**

Thank you!

smuckerb@miamioh.edu

http://www.users.miamioh.edu/smuckebj/

Twitter: @ByranSmucker