DEFINITIVE AUGMENTATION OF DEFINITIVE SCREENING DESIGNS, PART 1

PHIL KAY

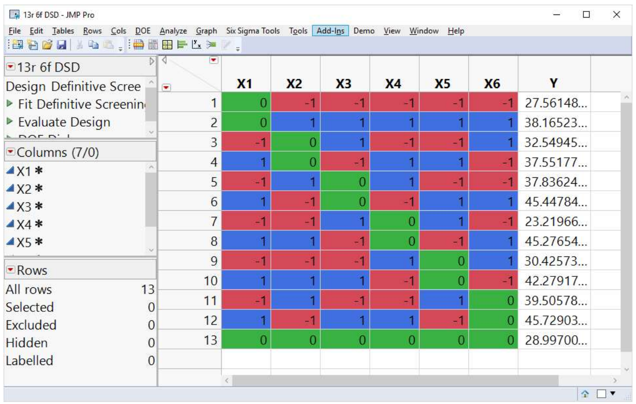
DISCOVERY SUMMIT, FRANKFURT, 15TH MARCH 2018

ABSTRACT

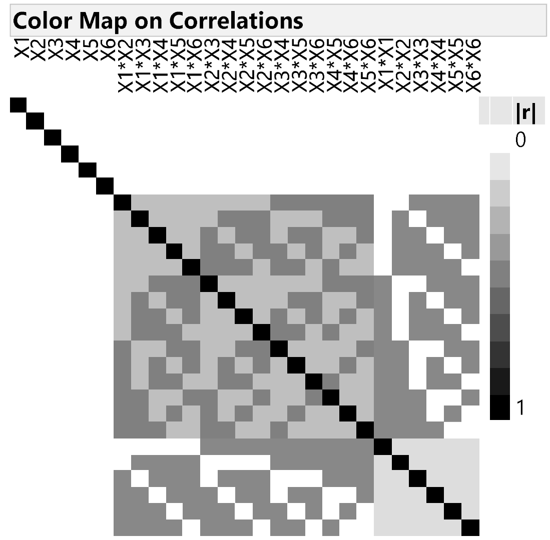
Definitive screening designs (DSDs) uniquely address the key needs of many experimenters. How else can we explain the rapid and enthusiastic adoption of DSDs since their discovery was published in 2011? For many experimenters, 13- or 17-run DSDs for five to seven factors are go-to designs when screening for the few driving factors. Along with 'Fit Definitive Screening' in JMP, you potentially have a simple, efficient and effective experimental workflow to find the important main effects, interactions and curvilinear behaviours of these factors. If only three of the factors are active, you can fit the full second-order RSM model and achieve screening and optimization in one step. But what if more than three factors are active? When ambiguity occurs is there a simple next step? Or does the complexity of this situation become a barrier to adoption of DSDs?

You will see simple ways to augment these DSDs, ensuring that the structure and properties can be preserved to maintain the benefits of DSDs. Consequently, more people in more situations can benefit from the workflow of sequential DOE and DSDs.

THE POWER OF DSDS

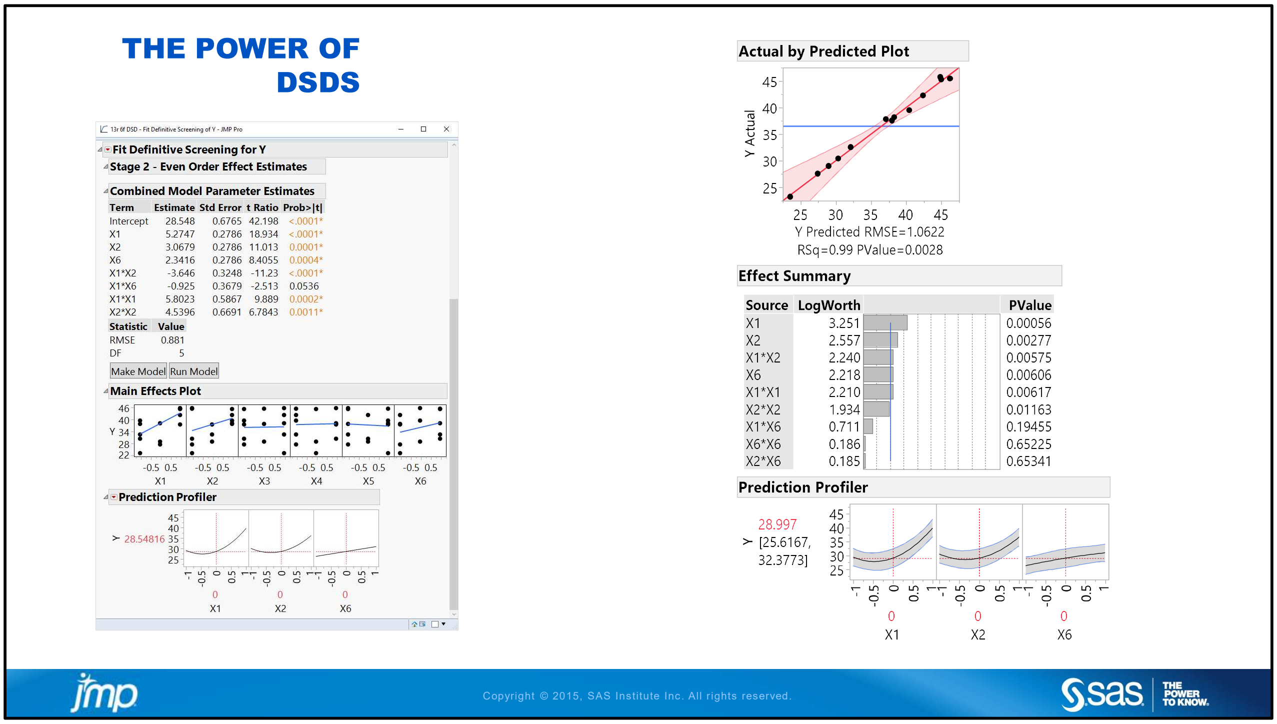


	X1	X2	X3	X4	X5	X6	Y
1	0	-1	-1	-1	-1	-1	27.56148...
2	0	1	1	1	1	1	38.16523...
3	-1	0	1	-1	-1	1	32.54945...
4	1	0	-1	1	1	-1	37.55177...
5	-1	1	0	1	-1	-1	37.83624...
6	1	-1	0	-1	1	1	45.44784...
7	-1	-1	1	0	1	-1	23.21966...
8	1	1	-1	0	-1	1	45.27654...
9	-1	-1	-1	1	0	1	30.42573...
10	1	1	1	-1	0	-1	42.27917...
11	-1	1	-1	-1	1	0	39.50578...
12	1	-1	1	1	-1	0	45.72903...
13	0	0	0	0	0	0	28.99700...

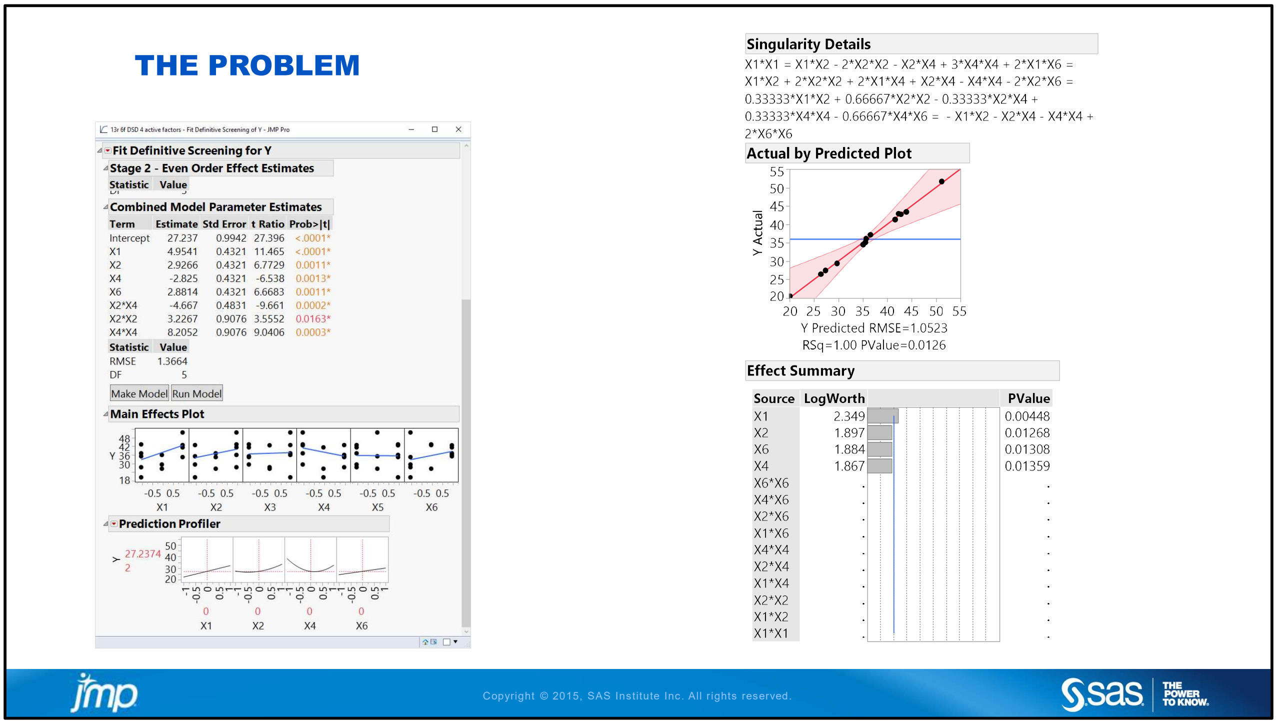


Small: n ~ 2k (n = number of runs, k = number of factors)
Main effects orthogonal vs each other
Main effects uncorrelated with all 2nd order effects
2-factor interactions not confounded with each other
Quadratic effects are estimable

The foldover pair structure is important for these properties

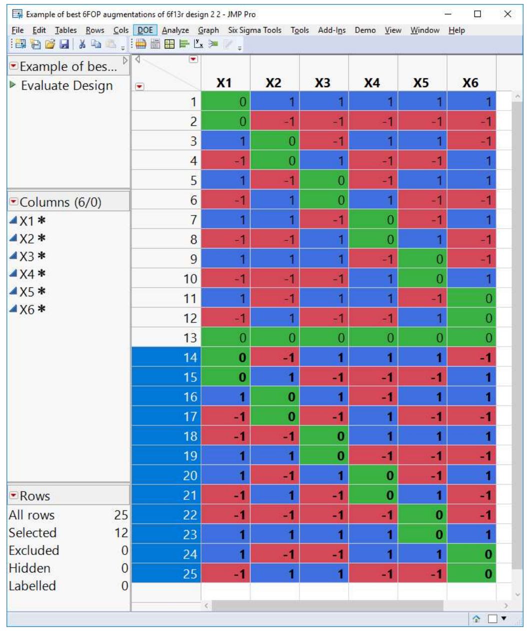


Quickly find active effects with Fit DSD
Can fit full RSM model for any 3 active factors
=> An effective workflow for many experimentalists in many situations



What if >3 factors are active?
DSD is good a detecting active main effects
But can't estimate full RSM
Low power to detect active 2nd order effects

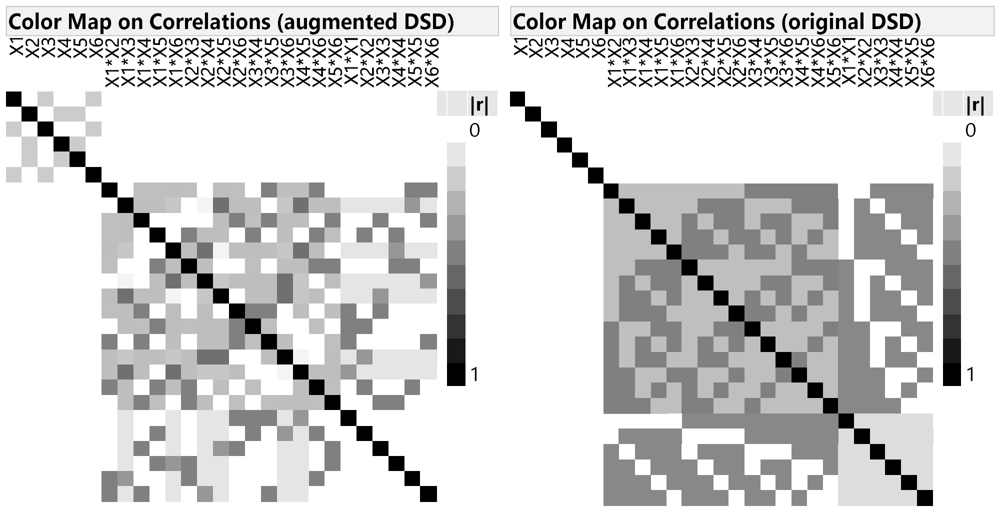
A SOLUTION



	X1	X2	X3	X4	X5	X6
1	0	1	1	1	1	1
2	0	-1	-1	-1	-1	-1
3	1	0	-1	1	1	-1
4	-1	0	1	-1	-1	1
5	1	-1	0	-1	1	1
6	-1	1	0	1	-1	-1
7	1	1	-1	0	-1	1
8	-1	-1	1	0	1	-1
9	1	1	1	-1	0	-1
10	-1	-1	-1	1	0	1
11	1	-1	1	1	-1	0
12	-1	1	-1	-1	1	0
13	0	0	0	0	0	0
14	0	-1	1	1	1	-1
15	0	1	-1	-1	-1	1
16	1	0	1	-1	1	1
17	-1	0	-1	1	-1	-1
18	-1	-1	0	1	1	1
19	1	1	0	-1	-1	-1
20	1	-1	1	0	-1	1
21	-1	1	-1	0	1	-1
22	-1	-1	-1	-1	0	-1
23	1	1	1	1	0	1
24	1	-1	-1	1	1	0
25	-1	1	1	-1	-1	0

12 run augmentation of the 13run 6factor DSD
Small(ish): n ~ 4k (n = number of runs, k = number of factors)
Main effects have 0 or 0.2 correlation vs each other
Main effects uncorrelated with all 2nd order effects
2-factor interactions not confounded with each other
Quadratic effects are estimable
Enables fitting of the full RSM for any *4* factors
You can still use Fit DSD for model selection

A SOLUTION



12 run augmentation of the 13run 6factor DSD
Small(ish): n ~ 4k (n = number of runs, k = number of factors)
Main effects have 0 or 0.2 correlation vs each other
Main effects uncorrelated with all 2nd order effects
2-factor interactions not confounded with each other
Quadratic effects are estimable
Enables fitting of the full RSM for any *4* factors
You can still use Fit DSD for model selection

HOW?

	Col1	Col2	Col3	Col4	Col5	Col6	Col7	Col8	Col9	Col10	Col11	Col12	Col13	Col14	Col15
1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
2	0	0	0	0	0	29.24	0	0	0	0	32.28	0	31.46	0	0
3	0	0	0	0	0	0	32.51	0	0	29.24	32.28	0	0	0	31.46
4	0	0	0	0	0	29.24	32.66	0	0	29.24	35.57	0	31.46	31.46	0
5	32.51	0	29.24	0	0	0	0	0	0	0	32.28	31.46	0	0	0
6	32.51	0	29.24	0	0	29.24	35.82	0	0	0	35.57	31.46	0	0	31.46
7	32.66	0	29.24	0	0	0	32.66	0	0	29.24	0	0	0	0	0
8	32.66	0	29.24	0	0	29.24	32.51	0	0	29.24	32.28	0	0	31.46	0
9	0	0	0	0	0	29.24	0	0	0	0	32.28	0	0	31.46	0
10	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
11	35.82	0	0	0	0	29.24	32.51	0	0	29.24	35.57	31.46	0	0	31.46
12	35.82	0	0	0	0	0	32.66	0	0	29.24	32.28	31.46	0	0	0
13	32.66	0	29.24	0	0	29.24	0	0	0	0	35.57	0	31.46	31.46	0

Constrained to augmentations with foldover pairs
-this enables Fit DSD
-and ensures MLEs are uncorrelated with 2nd order effects

The approach taken was to consider what are all the possibilities?
Then try them all to see which works best

Start with 6-factor 13-run DSD
Determine all possible unique augmentation runs
Determine all possible combinations of a number of these
Add first possible combination of runs and their fold-over "twins"
Compute D-efficiency for RSM model for all 4-factor (4f) projections
D-efficiency = 0 if model is not estimable
Code to repeat for all possible augmentation run combinations

And then tried this for different numbers of added runs
The result each time was a table

- 1 row for every augmentation possibility
- 1 column for each 4-factor projection (e.g. X1 X2 X3 X4 , X1 X2 X3 X5 …)
- Each cell is the D-efficiency for the full RSM model for that projection of that

augmentation
Then looking for rows (augmentations) with >0 in every column
i.e. The RSM model is estimable for every 4-factor projection

HOW?
SIZE OF THE CHALLENGE

X1	X2	X3	X4	X5	X6
-1	-1	-1	-1	-1	-1
-1	-1	-1	-1	-1	0
-1	-1	-1	-1	-1	1
-1	-1	-1	-1	0	-1
-1	-1	-1	-1	0	0
-1	-1	-1	-1	0	1
-1	-1	-1	-1	1	-1
-1	-1	-1	-1	1	0
-1	-1	-1	-1	1	1
-1	-1	-1	0	-1	-1
-1	-1	-1	0	-1	0
-1	-1	-1	0	-1	1
-1	-1	-1	0	0	-1
-1	-1	-1	0	0	0
-1	-1	-1	0	0	1
-1	-1	-1	0	1	-1
-1	-1	-1	0	1	0
-1	-1	-1	0	1	1
-1	-1	-1	1	-1	-1
-1	-1	-1	1	-1	0
-1	-1	-1	1	-1	1
-1	-1	-1	1	0	-1
-1	-1	-1	1	0	0
1	1	1	1	-1	0
1	1	1	1	-1	1
1	1	1	1	0	-1
1	1	1	1	0	0
1	1	1	1	0	1
1	1	1	1	1	-1
1	1	1	1	1	0
1	1	1	1	1	1

Too many possibilities!

For 6 factors at 3 levels (-1, 0, 1)
#Distinct runs = 3^6 = 729
=> 365 fold-over pairs (including 000000)

Adding 6 fold-over pairs as an augmentation
#Combinations = 365 choose 6 ≈3×〖10〗^12
15 4-factor projections
15×3×〖10〗^12 computations of D-efficiency

HOW? CONSTRAINING THE POSSIBILITIES

X1	X2	X3	X4	X5	X6
-1	-1	-1	-1	-1	-1
-1	-1	-1	-1	-1	0
-1	-1	-1	-1	-1	1
-1	-1	-1	-1	0	-1
-1	-1	-1	-1	0	0
-1	-1	-1	-1	0	1
-1	-1	-1	-1	1	-1
-1	-1	-1	-1	1	0
-1	-1	-1	-1	1	1
-1	-1	-1	0	-1	-1
-1	-1	-1	0	-1	0
-1	-1	-1	0	-1	1
-1	-1	-1	0	0	-1
-1	-1	-1	0	0	0
-1	-1	-1	0	0	1
-1	-1	-1	0	1	-1
-1	-1	-1	0	1	0
-1	-1	-1	0	1	1
-1	-1	-1	1	-1	-1
-1	-1	-1	1	-1	0
-1	-1	-1	1	-1	1
-1	-1	-1	1	0	-1
-1	-1	-1	1	0	0

Fold-over pair	X1	X2	X3	X4	X5	X6
1	0	±1	±1	±1	±1	±1
1	0	∓1	∓1	∓1	∓1	∓1
2	±1	0	±1	±1	±1	±1
2	∓1	0	∓1	∓1	∓1	∓1
3	±1	±1	0	±1	±1	±1
3	∓1	∓1	0	∓1	∓1	∓1
4	±1	±1	±1	0	±1	±1
4	∓1	∓1	∓1	0	∓1	∓1
5	±1	±1	±1	±1	0	±1
5	∓1	∓1	∓1	∓1	0	∓1
6	±1	±1	±1	±1	±1	0
6	∓1	∓1	∓1	∓1	∓1	0

jmp

sas THE POWER TO KNOW.

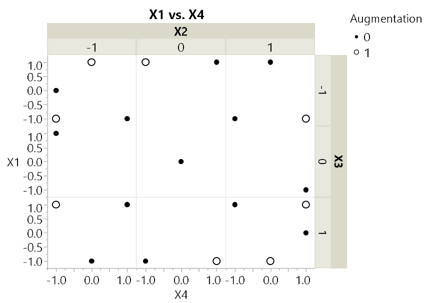
365 possible fold-over pairs

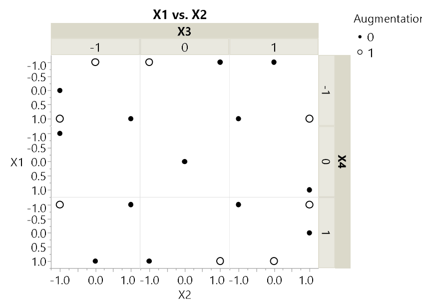
1. Consider only the 96 FOPs with 1 “0” per row
Still too many combinations
6FOPs: (96¦6) = 927,048,304
Days of computation
Can’t store results in laptop memory

So…
2. Consider only combinations with 1 “0” per column
6 factors, 16 distinct FOPs with the factor at 0
166 = 16,777,216 combinations
A few hours of computation
Found ~130,000 rows with no “0” – RSM estimable for every 4f projection
Out of the 130,000 we found 20 with the same high D-efficiency for the ME model for all 6 factors

Why did we constrain the probelm in this way?
Looking at smaller subsets of the problem pointed to these constraints
The constraints are consistent with the structure of the original design

OTHER SOLUTIONS: CONSIDER INDIVIDUAL PROJECTIONS





X1	X2	X3	X4
0	1	1	1
0	-1	-1	-1
1	0	-1	1
-1	0	1	-1
1	-1	0	-1
-1	1	0	1
1	1	-1	0
-1	-1	1	0
1	1	1	-1
-1	-1	-1	1
1	-1	1	1
-1	1	-1	-1
0	0	0	0
1	1	1	1
1	0	-1	-1
1	-1	1	-1
1	-1	-1	0
-1	-1	-1	-1
-1	0	1	1
-1	1	-1	1
-1	1	1	0

jmp

sas THE POWER TO KNOW.

Take each 4f projection of 6factor 13run DSD in turn

For 4 factors at 3 levels (-1, 0, 1)
#Distinct runs = 3^4 = 81
41 fold-over pairs (including 000000)
Adding k fold-over pairs (FOPs)
#Combinations = 41 choose k
820, 10660, 101270 (for k = 2, 3, 4)

Result:
RSM estimable with 4 FOPs
True for all 4f projections
Also…
Most efficient augmentations have no more than 1 “0” per run
All 4f projections (#1, #2 above) and “best” augmentn. are equivalent

Also considered the 17-run variation on the 6-factor DSD
Used same approach
Result
More complicated

Projections are not equivalent
RSM estimable with 2 addnl FOPs for some projections
Other projections require 3 addnl FOPs