

DOE Wizard – Alias-Optimal Designs



Revised: 12/11/2019



Summary	1
Example: Screening Design for 6 Factors in 12 Runs	2
Creating a Design with Minimal Aliasing	5
Evaluate Design	14
Adding a Third Level.....	16
Reducing Aliasing in Response Surface Designs	17
References.....	20

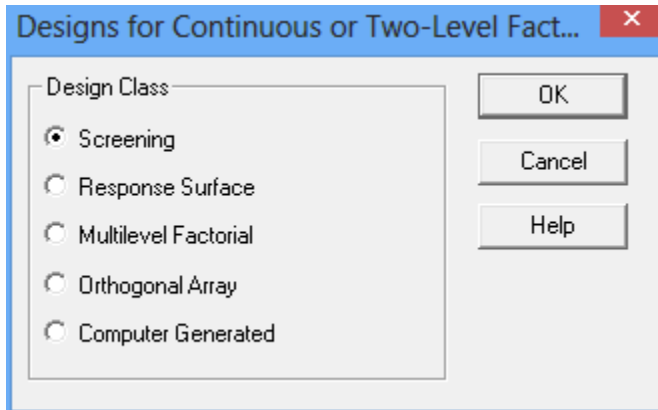
Summary

The *Alias-Optimal* designs generated by the DOE Wizard consider not only the precision in the estimated model coefficients but also potential bias in those estimates caused by active effects that are not in the assumed model. Criteria such as D-optimality do not take into account aliasing caused by omitted effects. Sometimes, alternative D-optimal designs may be subject to considerably different amounts of aliasing. At other times, a small reduction in the efficiency of the selected design may result in a large reduction in potential bias.

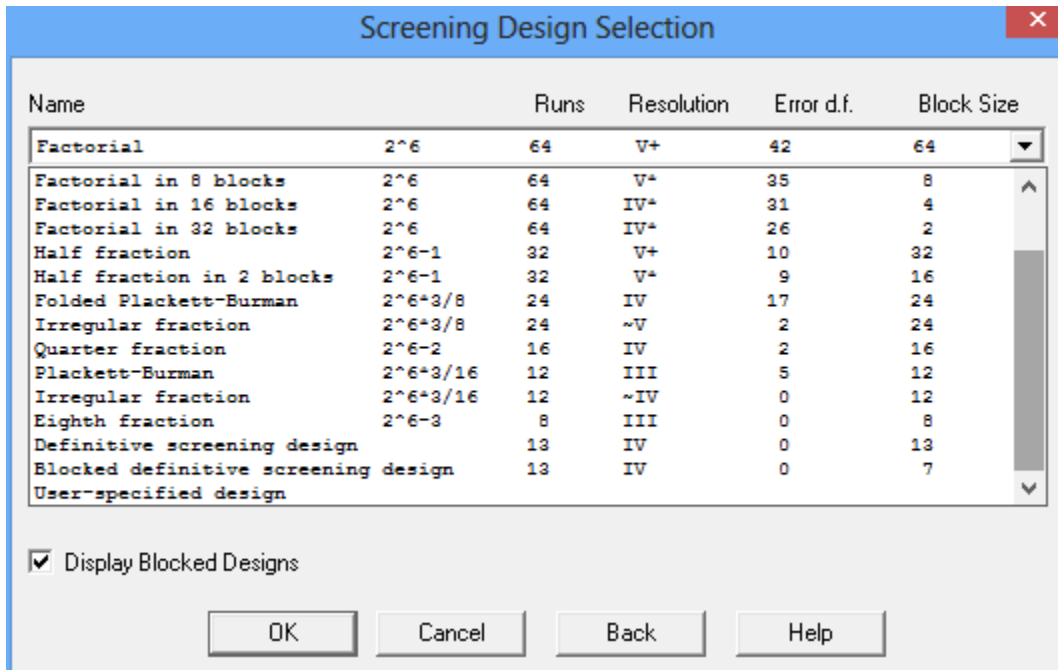
Jones and Nachtsheim (2011) present a procedure for constructing experimental designs with minimal aliasing. Given a selected statistical model and a maximum number of experimental runs that can be performed, they begin by constructing a D-optimal design for the specified model. Through a coordinate exchange algorithm, they then modify the selected model to maximize an objective function containing both the relative D-efficiency and the relative amount of aliasing associated with the design. The resulting designs give precise estimates of the fitted model coefficients while minimizing the amount of aliasing due to effects that are not in the primary model.

Example: Screening Design for 6 Factors in 12 Runs

Suppose that an experimenter wished to determine which of 6 factors have the largest effect on a response. The Statgraphics *DOE Wizard* provides a step-by-step procedure for constructing an experimental design for such purposes. During the third step of that process, the following dialog box is displayed:



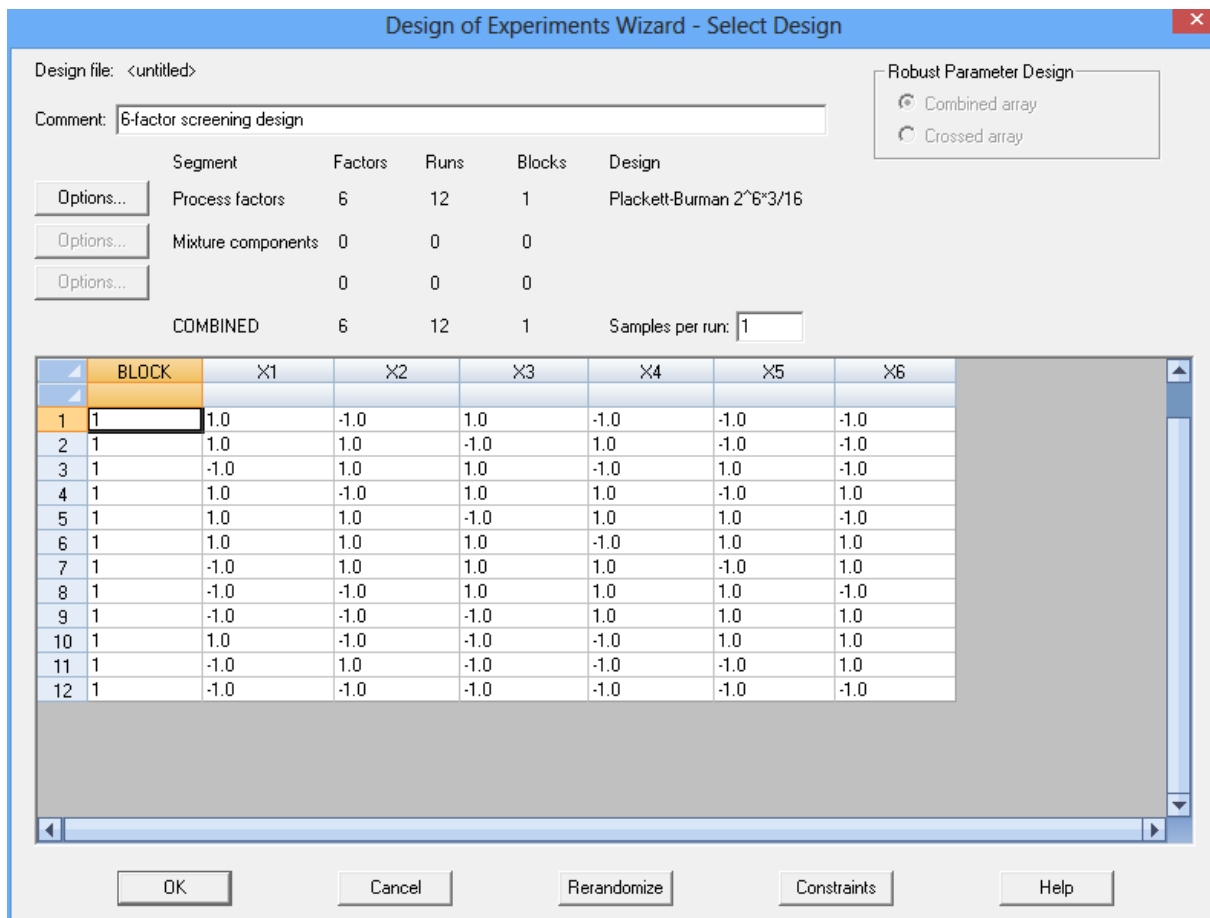
If screening is selected, the following list of designs will be offered:



A common choice for screening 6 factors is the 12-run Plackett-Burman design, which is capable of fitting the first-order model

$$Y = \beta_0 + \beta_1X_1 + \beta_2X_2 + \beta_3X_3 + \beta_4X_4 + \beta_5X_5 + \beta_6X_6$$

Since the model contains 7 coefficients, there will be 5 degrees of freedom available to estimate the experimental error. The design matrix X consists of 12 rows and 7 columns, as shown below:



The ability of a screening design to estimate the coefficients of a model is often characterized by its D-efficiency

$$D = 100 \frac{|X'X|^{1/p}}{n} \tag{1}$$

where p is the number of coefficients in the model and n is the number of experimental runs. The Plackett-Burman design is “D-optimal”, meaning that it has the highest D-efficiency of any set of 12 runs for the specified model.

An interesting way to examine the properties of any design is to display the correlations between the estimated coefficients when the model is fit. The correlation matrix for the Plackett-Burman design is shown below:

Correlation Matrix

	A	B	C	D	E	F
A	1.0000	0.0000	0.0000	0.0000	0.0000	0.0000
B	0.0000	1.0000	0.0000	0.0000	0.0000	0.0000
C	0.0000	0.0000	1.0000	0.0000	0.0000	0.0000
D	0.0000	0.0000	0.0000	1.0000	0.0000	0.0000
E	0.0000	0.0000	0.0000	0.0000	1.0000	0.0000
F	0.0000	0.0000	0.0000	0.0000	0.0000	1.0000

In this case, all of the main effects are orthogonal to each other implying that the standard errors of the coefficients will not be inflated.

While high D-efficiency is a desirable property, it says nothing about how terms that are not included in the model affect the estimated coefficients. For example, there are 15 possible 2-factor interactions of the form $X_i X_j$. To examine their influence on the estimated coefficients, a second matrix X_A could be constructed consisting of 12 rows and 15 columns, where each column contains the crossproduct of 2 columns in the X matrix. An alias matrix could then be constructed of the form

$$A = (X'X)^{-1}X'X_A \quad (2)$$

For the Plackett-Burman design, A has 7 rows and 15 columns as shown below (values equal to 0 are omitted):

Alias Matrix

Effect	AB	AC	AD	AE	AF	BC	BD	BE	BF	CD	CE	CF	DE	DF	EF
constant															
A						-0.333	0.333	0.333	-0.333	-0.333	-0.333	0.333	-0.333	-0.333	0.333
B		-0.333	0.333	0.333	-0.333					-0.333	0.333	0.333	-0.333	-0.333	-0.333
C	-0.333		-0.333	-0.333	0.333		-0.333	0.333	0.333				-0.333	0.333	-0.333
D	0.333	-0.333		-0.333	-0.333	-0.333		-0.333	-0.333		-0.333	0.333			-0.333
E	0.333	-0.333	-0.333		0.333	0.333	-0.333		-0.333	-0.333		-0.333			-0.333
F	-0.333	0.333	-0.333	0.333		0.333	-0.333	-0.333		0.333	-0.333		-0.333		

A non-zero entry in a cell implies that the interaction in that column will have an effect on the coefficient in that row. In this design, each interaction affects all main effects except the 2 that it involves.

A common measure of the potential for bias is the sum of squares of the entries in the alias matrix, which may be calculated as $\text{Trace}(A'A)$. For the 12-run Plackett-Burman design, $\text{Trace}(A'A) = 60/9 = 6.6667$.

Designs with Minimal Aliasing

Jones and Nachtsheim (2011) suggested that when constructing a screening design, both the D-efficiency and the alias matrix should be taken into account. In some cases, alternative D-optimal designs exist which have the same D-efficiency as another design but less aliasing. In other cases, a small decrease in D-efficiency may result in a large decrease in the amount of bias due

to aliasing. They describe a method for finding an experimental design which minimizes $Trace(A'A)$ while insuring that the design has a D-efficiency of at least $\lambda_D\%$ of the D-optimal design. The methodology involves Lagrange multipliers and a coordinate exchange algorithm and is fully explained in their paper.

The next section shows how to use Statgraphics to construct a screening design with minimal aliasing.

Creating a Design with Minimal Aliasing

To begin the design creation process, start with an empty StatFolio. Select *DOE – Experimental Design Wizard* to load the DOE Wizard’s main window. Then push each button in sequence to create the design.

Step #1 – Define Responses

The first step of the design creation process displays a dialog box used to specify the response variables. For illustration purposes, a single response variable will suffice:



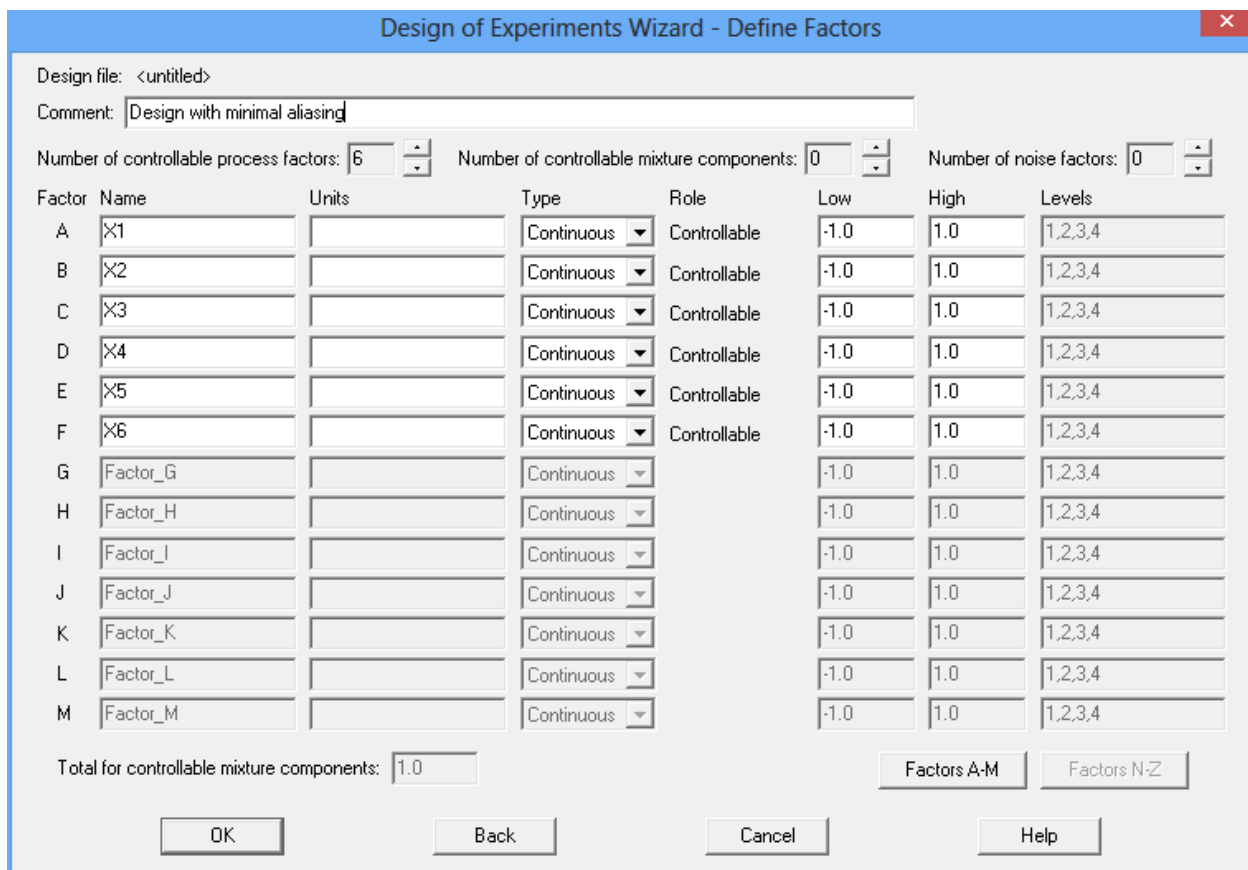
Response	Name	Units	Analyze	Goal	Target	Impact (1-5)	Sensitivity	Minimum	Maximum
1	Y		Mean	Maximize	0.5	3.0	Medium		
2	Var_2		Mean	Maximize	0.5	3.0	Medium		
3	Var_3		Mean	Maximize	0.5	3.0	Medium		
4	Var_4		Mean	Maximize	0.5	3.0	Medium		
5	Var_5		Mean	Maximize	0.5	3.0	Medium		
6	Var_6		Mean	Maximize	0.5	3.0	Medium		
7	Var_7		Mean	Maximize	0.5	3.0	Medium		
8	Var_8		Mean	Maximize	0.5	3.0	Medium		
9	Var_9		Mean	Maximize	0.5	3.0	Medium		
10	Var_10		Mean	Maximize	0.5	3.0	Medium		
11	Var_11		Mean	Maximize	0.5	3.0	Medium		
12	Var_12		Mean	Maximize	0.5	3.0	Medium		
13	Var_13		Mean	Maximize	0.5	3.0	Medium		
14	Var_14		Mean	Maximize	0.5	3.0	Medium		
15	Var_15		Mean	Maximize	0.5	3.0	Medium		
16	Var_16		Mean	Maximize	0.5	3.0	Medium		

- **Name:** The name for the response.
- **Units:** The units of the response.

- **Analyze:** The parameter of interest (usually *mean*).
- **Goal:** The goal of the experiment (maximize the response, minimize the response, or hit a target value).
- **Impact:** The relative importance of each response if performing multi-response optimization.
- **Sensitivity:** The importance of being close to the best desired value.
- **Minimum and Maximum:** The range of acceptable response values, if any.

Step #2 – Define Experimental Factors

The second step displays a dialog box on which to specify the factors that will be varied:



Design file: <untitled>
 Comment: Design with minimal aliasing

Number of controllable process factors: 6 Number of controllable mixture components: 0 Number of noise factors: 0

Factor	Name	Units	Type	Role	Low	High	Levels
A	X1		Continuous	Controllable	-1.0	1.0	1,2,3,4
B	X2		Continuous	Controllable	-1.0	1.0	1,2,3,4
C	X3		Continuous	Controllable	-1.0	1.0	1,2,3,4
D	X4		Continuous	Controllable	-1.0	1.0	1,2,3,4
E	X5		Continuous	Controllable	-1.0	1.0	1,2,3,4
F	X6		Continuous	Controllable	-1.0	1.0	1,2,3,4
G	Factor_G		Continuous		-1.0	1.0	1,2,3,4
H	Factor_H		Continuous		-1.0	1.0	1,2,3,4
I	Factor_I		Continuous		-1.0	1.0	1,2,3,4
J	Factor_J		Continuous		-1.0	1.0	1,2,3,4
K	Factor_K		Continuous		-1.0	1.0	1,2,3,4
L	Factor_L		Continuous		-1.0	1.0	1,2,3,4
M	Factor_M		Continuous		-1.0	1.0	1,2,3,4

Total for controllable mixture components: 1.0 Factors A-M Factors N-Z

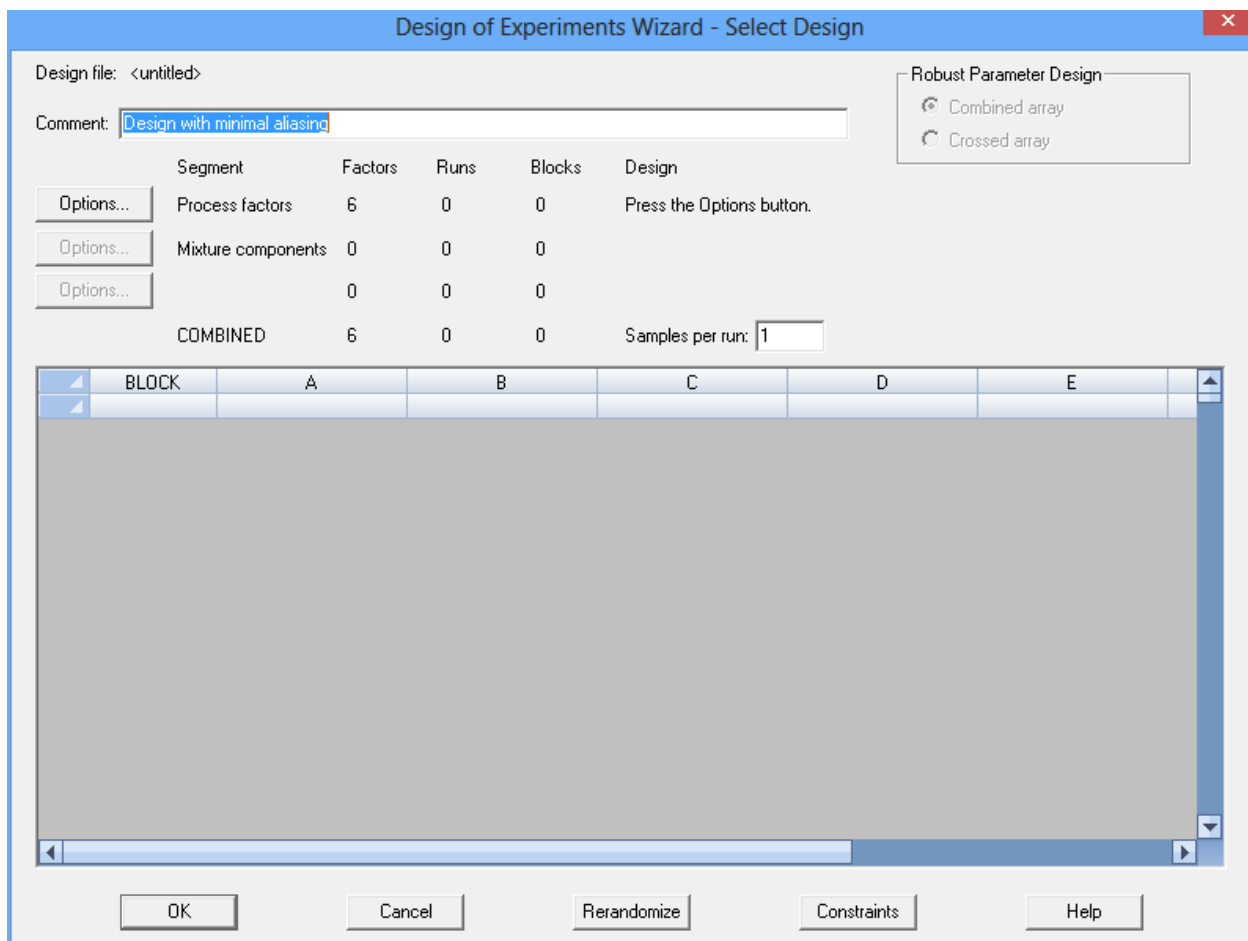
OK Back Cancel Help

- **Name** – Each factor must be assigned a unique name.
- **Units** – Units are optional.
- **Type** – The type of factor. Alias-optimal designs are only available if all factors are either continuous or categorical with 2 levels.

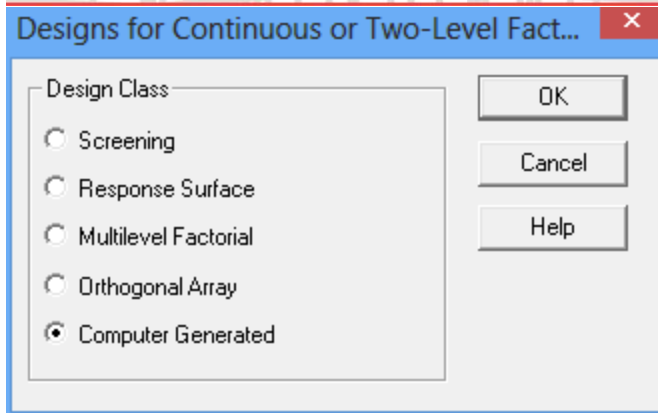
- **Role** – Alias-optimal designs are only available when all factors are controllable process factors.
- **Low** - the lower level L_j for a continuous factors
- **High** - the upper level U_j for a continuous factor.
- **Levels** – the allowable levels of a categorical factor.

Step #3 - Select Design

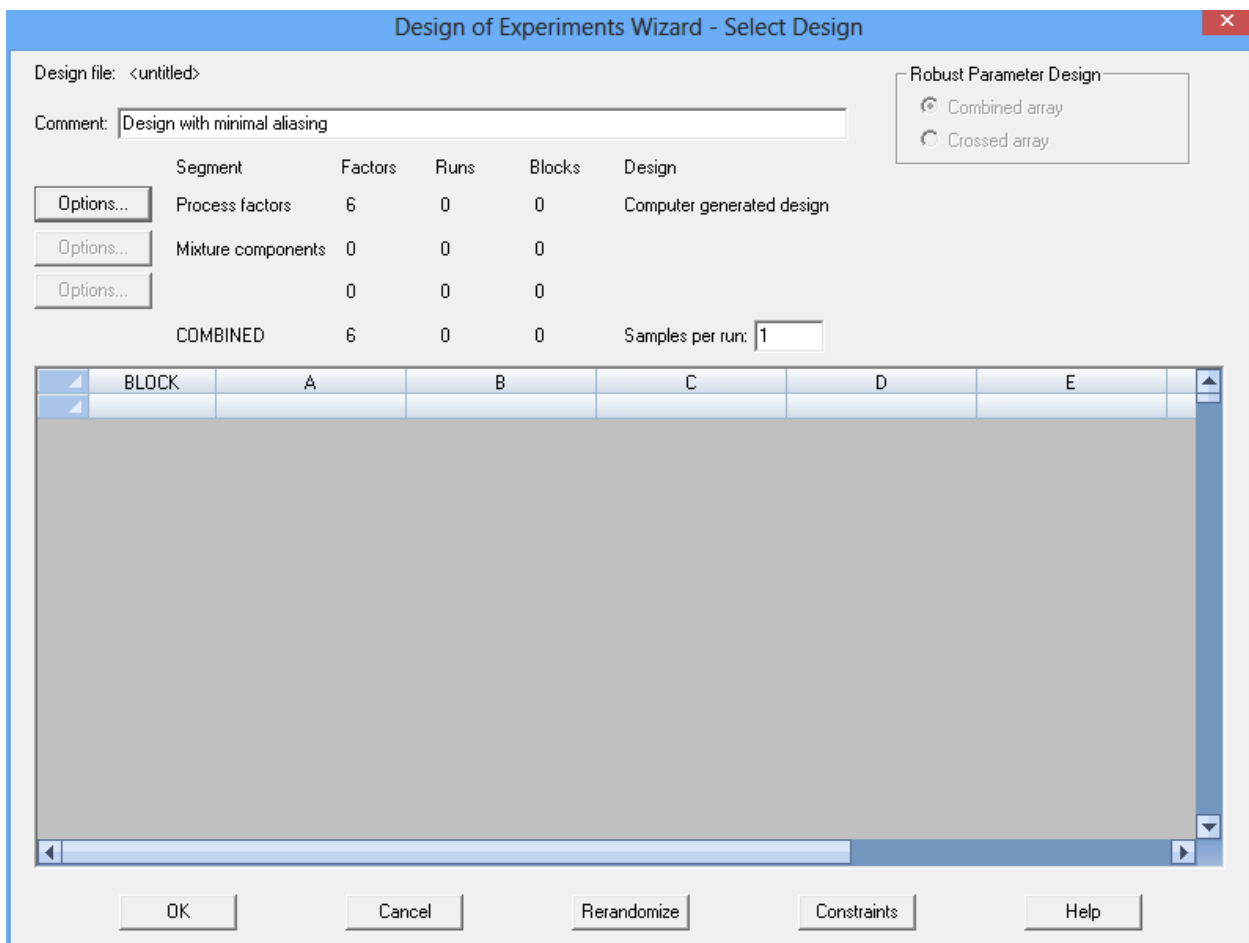
The third step begins by displaying the dialog box shown below:



Since all of the factors are controllable process factors, only one *Options* button is enabled. Pressing that button displays a second dialog box:



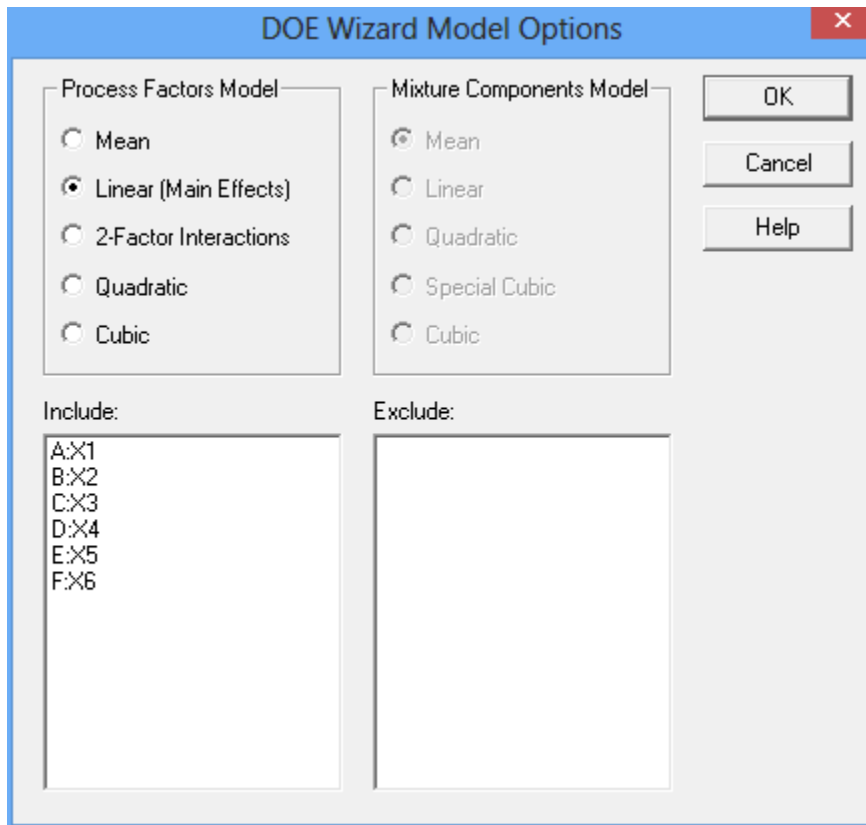
Select *Computer Generated* and press *OK*, which will return you to the *Select Design* dialog box:



Press *OK* to save the design selection and return to the DOE Wizard's main window.

Step #4: Specify Model

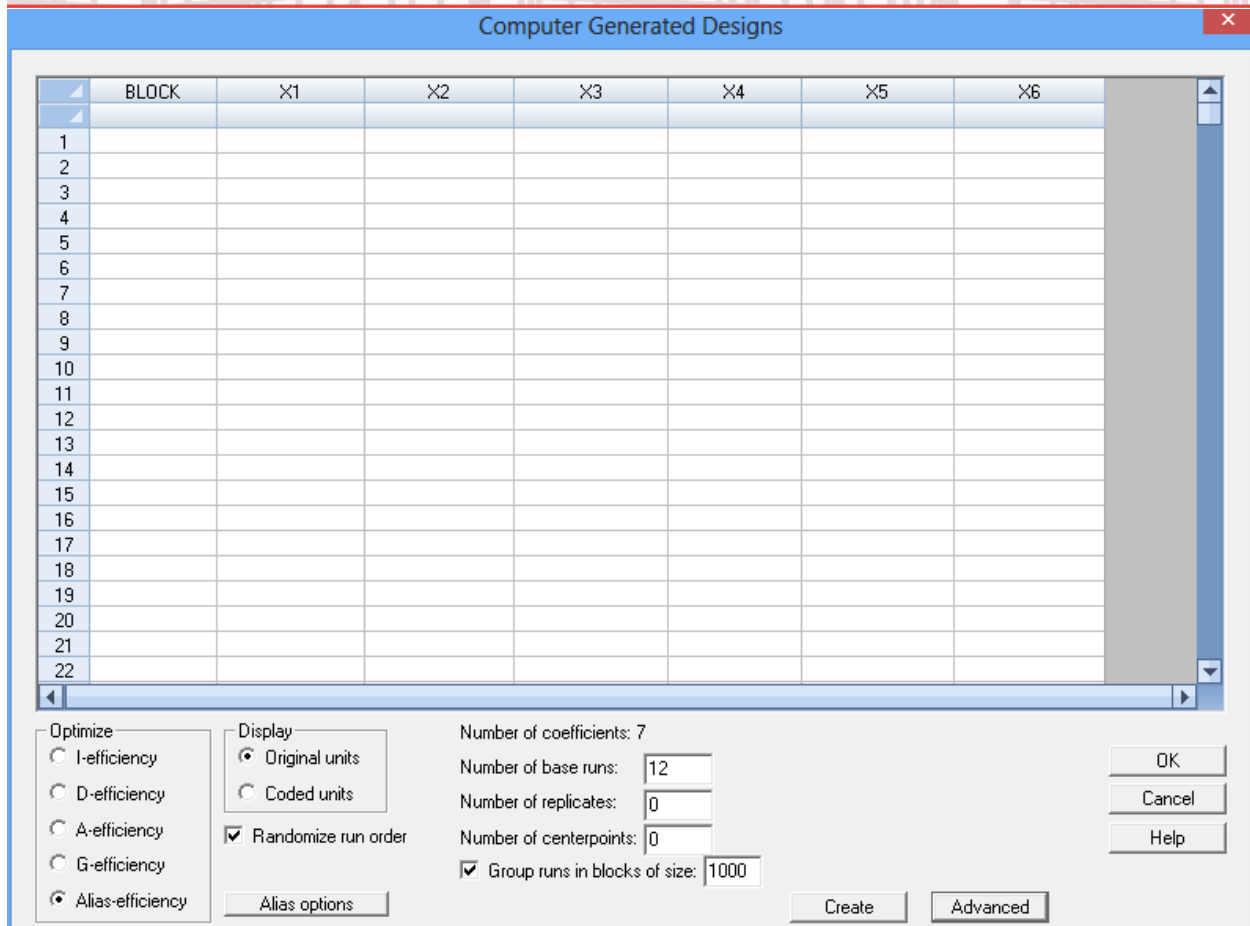
The next step in the design selection process specifies the primary model that will be fit to the response data. Pressing the fourth button on the DOE Wizard’s toolbar displays a dialog box to make that choice:



In this example, only the 6 main effects will be estimated. In other cases, additional effects could be specified.

Step #5: Select runs

The next step selects the runs to be performed. Press the *Step 5: Select Runs* button on the DOE Wizard toolbar to display the following dialog box:

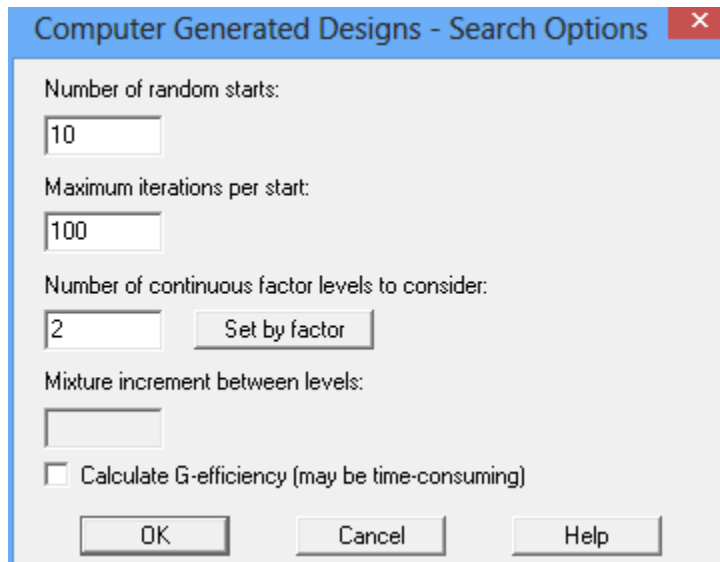


The dialog box shows the number of coefficients in the model to be estimated. At least as many runs must be selected as there are coefficients. In this example, the experimenters decided to perform 12 runs.

There are several important fields to be completed:

- *Optimize* - the criterion to be used to select the experimental runs. Set it to “alias-efficiency” to create an alias-optimal design.
- *Display* - whether the runs should be displayed in their original units or coded units.
- *Randomize run order* - whether the order of the runs should be randomized.
- *Number of base runs* - the number of different combinations of the factors that should be generated. This number must be greater than or equal to the number of coefficients in the model that will be fit to the data.
- *Number of replicates* - the number of additional runs to be added that are replicates of one or more of the base runs.

- *Number of centerpoints* - the number of additional runs to be added at the center of the experimental region.
- *Group runs in blocks of size* – If desired, the new runs may be placed in blocks of the specified size. Additional terms will be added to the model to allow for differences between the blocks. Selecting this option also places any experimental runs that have already been performed in a different block than the new runs.
- *Advanced* button – displays a dialog box for changing the search options:



Number of random starts – number of times the algorithm will search for a D-optimal design from a different random start. After a D-optimal design is selected, the algorithm then looks for a better design that reduces the amount of aliasing.

Maximum iterations per start – maximum number of times that the algorithm will try exchanging runs before a solution is accepted.

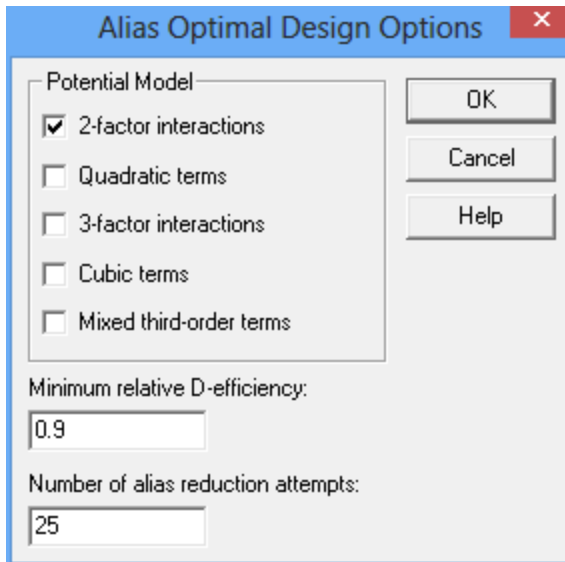
Number of factor levels to consider - number of levels at which experimental runs may be performed for continuous factors, ranging from the low level of the factor to the high level. Specifying a larger number increases the time required to create the design but may improve the design efficiency. If different numbers of levels are desired for different factors, they may be set by pressing the *Set by factor* button. For the current example, set the number to 2 so that only low and high levels of the factors will be considered.

Mixture increment between levels – for experiments containing mixture components, the amount by which each component will be changed when attempting to find an optimal set of runs.

Calculate G-efficiency - whether the G-efficiency of the design should be calculated and displayed. This statistic requires calculating the prediction variance at every candidate

point considered when constructing the design, which can be very large if there are many factors or factor levels to consider.

- *Alias options button* – displays an additional dialog box with options specific to alias-optimal designs:



Potential Model – the terms to be included in the alias matrix A . 2-factor interactions are terms such as X_1X_2 . Quadratic terms are terms such as X_1^2 . 3-factor interactions are terms such as $X_1X_2X_3$. Cubic terms are terms such as X_1^3 . Mixed third-order terms are terms such as $X_1^2X_2$.

Minimum relative D-efficiency – λ_D , the smallest allowable ratio of the resulting D-efficiency to that of a D-optimal design.

Number of alias reduction attempts – to increase the chance of obtaining a global optimum rather than a local optimum, the design creation process is repeated the indicated number of times. The result returned is the best design among those found.

To generate the design, press the *Create* button. When the algorithm is complete, the selected experimental runs will be added to the dialog box:

Computer Generated Designs

	BLOCK	X1	X2	X3	X4	X5	X6
1	1	1.0	1.0	1.0	-1.0	1.0	1.0
2	1	1.0	-1.0	-1.0	-1.0	1.0	1.0
3	1	1.0	1.0	-1.0	1.0	1.0	-1.0
4	1	1.0	1.0	-1.0	1.0	-1.0	1.0
5	1	-1.0	1.0	-1.0	1.0	1.0	1.0
6	1	1.0	-1.0	1.0	1.0	1.0	1.0
7	1	-1.0	-1.0	1.0	-1.0	-1.0	1.0
8	1	-1.0	1.0	-1.0	-1.0	-1.0	-1.0
9	1	-1.0	-1.0	-1.0	1.0	-1.0	-1.0
10	1	1.0	-1.0	1.0	-1.0	-1.0	-1.0
11	1	-1.0	1.0	1.0	1.0	-1.0	-1.0
12	1	-1.0	-1.0	1.0	-1.0	1.0	-1.0
13							
14							
15							
16							
17							
18							
19							
20							
21							
22							

Optimize

I-efficiency

D-efficiency

A-efficiency

G-efficiency

Alias-efficiency

Display

Original units

Coded units

Randomize run order

Number of coefficients: 7

Number of base runs:

Number of replicates:

Number of centerpoints:

Group runs in blocks of size:

Average prediction variance: 0.283333

D-efficiency: 91.78%

A-efficiency: 85.37%

trace(A'A): 0.666667

The design obtained has a D-efficiency of 91.78% compared to 100% for the Plackett-Burman design, which satisfies the criteria specified of $\lambda_D = 0.9$. The amount of bias due to aliasing as measured by Trace(A'A) has dropped dramatically and is only 10% of that calculated for the Plackett-Burman design.

To save the design, press OK. The main DOE Wizard window will now reflect the selected design:

Experimental Design Wizard

Step 1: Define responses |
 Step 3: Select design |
 Step 5: Select runs |
 Step 7: Save experiment |
 Step 9: Optimize responses |
 Step 11: Augment design
Step 2: Define exp. factors |
 Step 4: Specify model |
 Step 6: Evaluate design |
 Step 8: Analyze data |
 Step 10: Save results |
 Step 12: Extrapolate

Experimental Design Wizard

Step 1: Define the response variables to be measured

Name	Units	Analyze	Goal	Target	Impact	Sensitivity	Low	High
Y		Mean	Maximize		3.0	Medium		

Step 2: Define the experimental factors to be varied

Name	Units	Type	Role	Low	High	Levels
A:A		Continuous	Controllable	-1.0	1.0	
B:B		Continuous	Controllable	-1.0	1.0	
C:C		Continuous	Controllable	-1.0	1.0	
D:D		Continuous	Controllable	-1.0	1.0	
E:E		Continuous	Controllable	-1.0	1.0	
F:F		Continuous	Controllable	-1.0	1.0	

Step 3: Select the experimental design

Type of	Design	Centerpoints	Centerpoint	Design is	Number of	Total	Total	Error
Factors	Type	Per Block	Placement	Randomized	Replicates	Runs	Blocks	D.F.
Process	Computer generated design							

Number of samples per run: 1

Step 4: Specify the initial model to be fit to the experimental results

Factors	Model	Coefficients	Excluded effects
Process	linear (main effects)	7	

Step 5: Select an optimal subset of the runs (optional)
12 runs selected

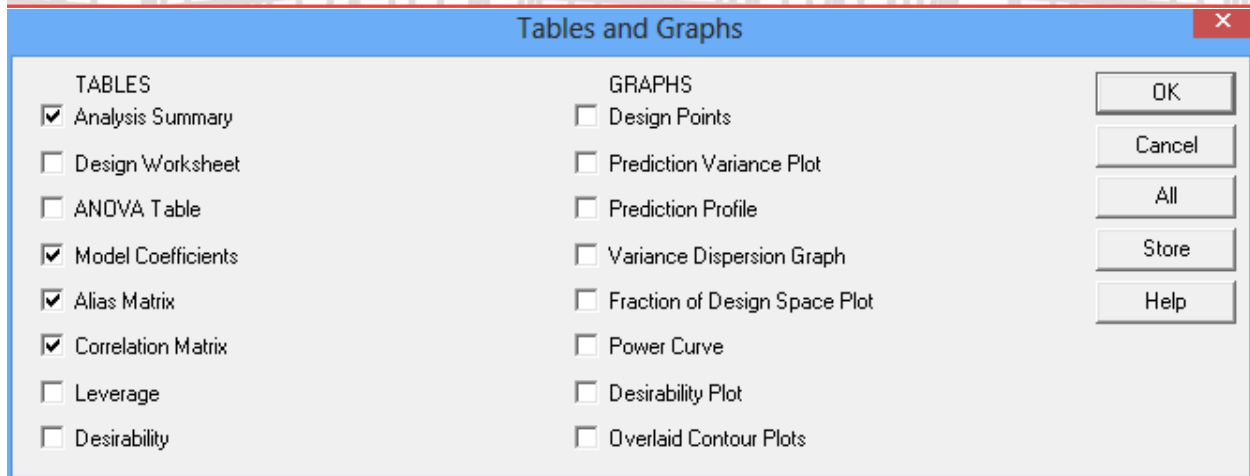
Step 6: Select tables and graphs to evaluate the selected runs
To display design diagnostics, use the checkbox on the Analysis Options dialog box.

Step 7: Save the experiment
Design file: <untitled> (Design with minimal aliasing)

If the selection is acceptable, press *Step 7: Save experiment* to save the design.

Evaluate Design

After the design has been created, press the button labeled *Step 6: Evaluate Design* on the DOE Wizard toolbar to display various design diagnostics:



The *Model Coefficients* table shows the relative standard error of each coefficient in the model to be estimated:

Model Coefficients						
				<i>Power at</i>	<i>Power at</i>	<i>Power at</i>
<i>Coefficient</i>	<i>Standard Error</i>	<i>VIF</i>	<i>Ri-Squared</i>	<i>SN = 0.5</i>	<i>SN = 1.0</i>	<i>SN = 2.0</i>
constant	0.288675			10.97%	29.15%	78.91%
A	0.316228	1.2	0.166667	9.96%	25.18%	71.61%
B	0.316228	1.2	0.166667	9.96%	25.18%	71.61%
C	0.316228	1.2	0.166667	9.96%	25.18%	71.61%
D	0.316228	1.2	0.166667	9.96%	25.18%	71.61%
E	0.316228	1.2	0.166667	9.96%	25.18%	71.61%
F	0.316228	1.2	0.166667	9.96%	25.18%	71.61%

alpha = 5.0%, sigma estimated from total error with 5 d.f.

The standard error is relative in the sense that it is the multiple of the residual standard error, which is not known until the experiment has been performed. Of particular interest are the VIFs (Variance Inflation Factors), which show how much the variance of each coefficient has been increased relative to a perfectly orthogonal design. Since the largest VIF is 1.2, there has been relatively little variance inflation.

The *Alias Matrix* shows that the bias due to any potential 2-factor interactions is now limited to the constant term:

Alias Matrix															
Effect	AB	AC	AD	AE	AF	BC	BD	BE	BF	CD	CE	CF	DE	DF	EF
constant			0.3333	0.3333		-0.3333			-0.3333			0.3333	0.3333		
A															
B															
C															
D															
E															
F															

None of the main effects will be corrupted by any out-of-the-model 2-factor interaction.

The *Correlation Matrix* shows the correlation amongst the estimated model coefficients:

	A	B	C	D	E	F
A	1.0000	0.0000	0.0000	0.3333	0.3333	0.0000
B	0.0000	1.0000	-0.3333	0.0000	0.0000	-0.3333
C	0.0000	-0.3333	1.0000	0.0000	0.0000	0.3333
D	0.3333	0.0000	0.0000	1.0000	0.3333	0.0000
E	0.3333	0.0000	0.0000	0.3333	1.0000	0.0000
F	0.0000	-0.3333	0.3333	0.0000	0.0000	1.0000

As a tradeoff for the reduction in bias due to aliasing, there is now some correlation amongst the main effects. However, that correlation is relatively small.

Adding a Third Level

The resulting minimal aliasing design depends on several factors:

1. The number of levels considered for each factor.
2. The percentage reduction allowed in D-efficiency.

For example, if λ_D is reduced to 0.8 and 3 levels of each factor are allowed, the resulting design is shown below:

Computer Generated Designs

	BLOCK	X1	X2	X3	X4	X5	X6
1	1	1.0	-1.0	-1.0	-1.0	0.0	1.0
2	1	1.0	0.0	-1.0	1.0	1.0	-1.0
3	1	-1.0	-1.0	0.0	-1.0	1.0	-1.0
4	1	-1.0	1.0	1.0	1.0	0.0	-1.0
5	1	1.0	-1.0	1.0	0.0	-1.0	-1.0
6	1	0.0	1.0	-1.0	-1.0	-1.0	-1.0
7	1	-1.0	-1.0	-1.0	1.0	-1.0	0.0
8	1	1.0	1.0	1.0	-1.0	1.0	0.0
9	1	1.0	1.0	0.0	1.0	-1.0	1.0
10	1	-1.0	1.0	-1.0	0.0	1.0	1.0
11	1	0.0	-1.0	1.0	1.0	1.0	1.0
12	1	-1.0	0.0	1.0	-1.0	-1.0	1.0
13							
14							
15							
16							
17							
18							
19							
20							
21							
22							

Optimize

I-efficiency

D-efficiency

A-efficiency

G-efficiency

Alias-efficiency

Display

Original units

Coded units

Randomize run order

Alias options

Number of coefficients: 7

Number of base runs:

Number of replicates:

Number of centerpoints:

Group runs in blocks of size:

Average prediction variance: 0.283333

D-efficiency: 85.53%

A-efficiency: 85.37%

trace(AA): 0.0

OK

Cancel

Help

Create

Advanced

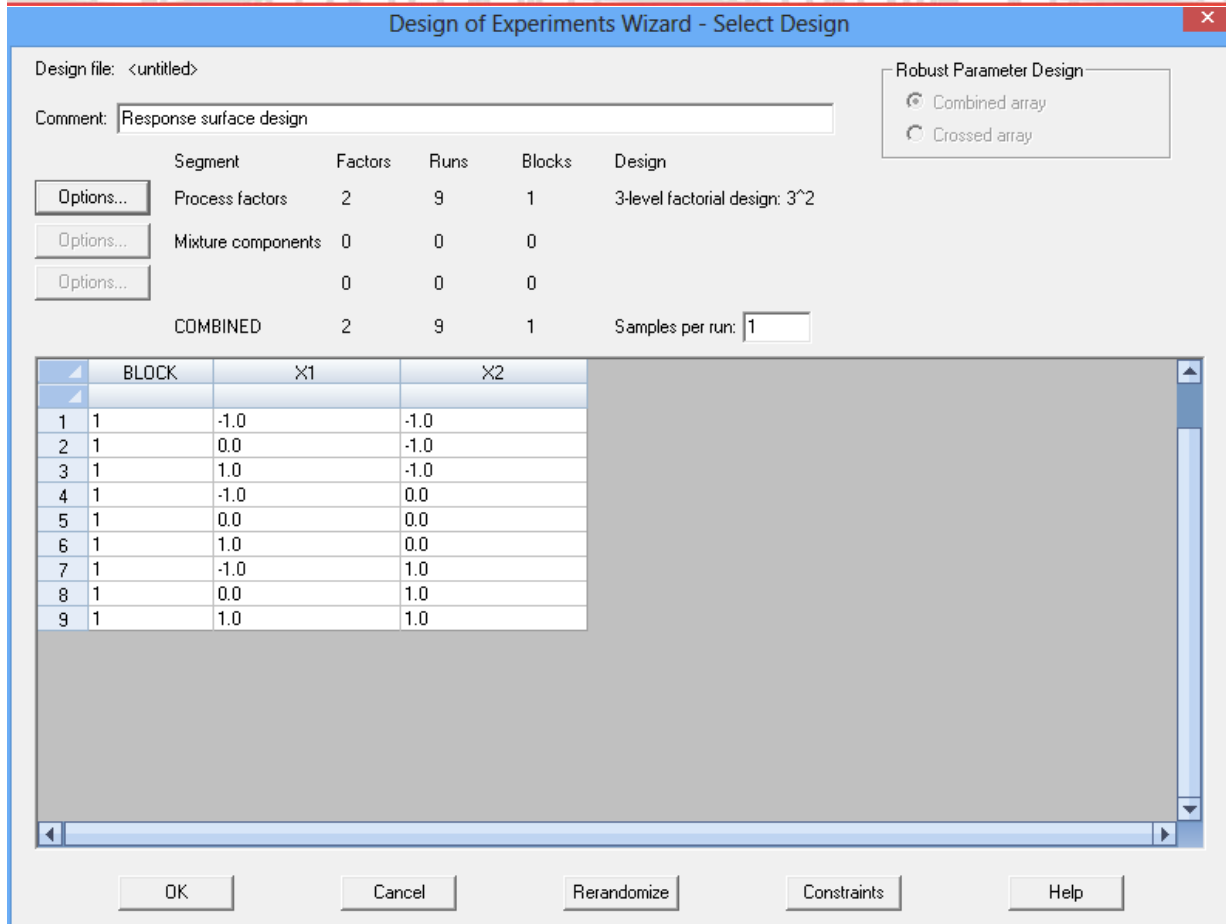
It turns out that all bias due to aliasing from 2-factor interactions disappears by running each of the 6 factors twice at a value halfway between its low and high levels. The D-efficiency falls to 85.53%, but the average prediction variance is unchanged.

Reducing Aliasing in Response Surface Designs

In the case of a response surface design, the primary model normally contains all 2-factor interactions and quadratic terms. The potential model then contains cubic terms similar to X_1^3 and $X_1^2X_2$. For example, consider mapping a response surface for 2 factors using an experimental design with 9 runs. The D-optimal design for the model

$$Y = \beta_0 + \beta_1X_1 + \beta_2X_2 + \beta_3X_1^2 + \beta_4X_2^2 + \beta_5X_1X_2 \tag{3}$$

is the 3 x 3 full factorial design shown below:



Note the orthogonality in the correlation matrix:

Correlation Matrix

	A	B	AA	AB	BB
A	1.0000	0.0000	0.0000	0.0000	0.0000
B	0.0000	1.0000	0.0000	0.0000	0.0000
AA	0.0000	0.0000	1.0000	0.0000	0.0000
AB	0.0000	0.0000	0.0000	1.0000	0.0000
BB	0.0000	0.0000	0.0000	0.0000	1.0000

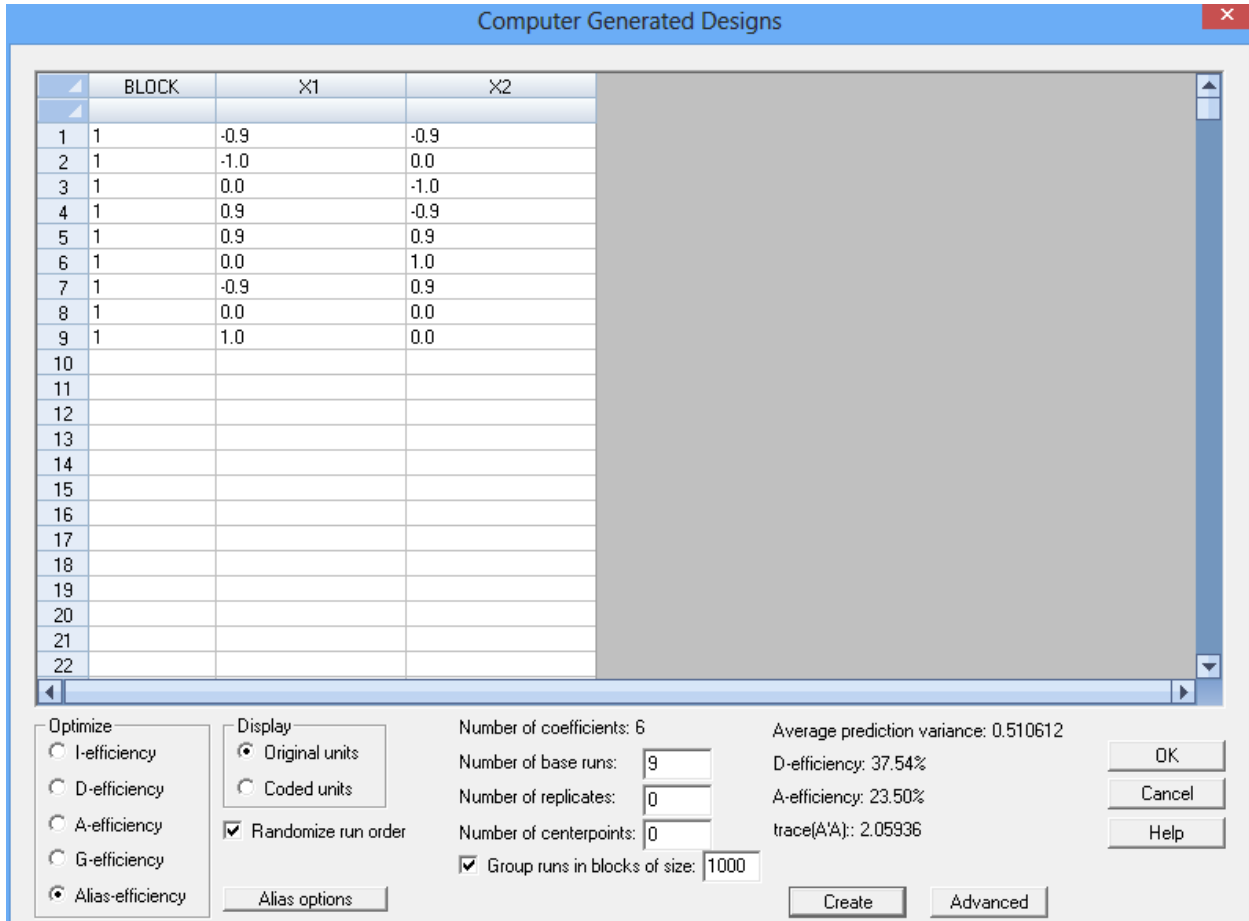
On the other hand, there is serious bias in the main effects caused by potential third order terms as shown in the alias matrix:

Alias Matrix

Effect	AAA	AAB	ABB	BBB
constant				
A	1.0000		0.6667	
B		0.6667		1.0000
AA				
AB				
BB				

Note that for this design, $\text{Trace}(A'A) = 3.3333$.

To generate a design with less aliasing, we will let $\lambda_D = 0.81$ and allow the algorithm to try 21 values of each factor evenly spaced between -1 and +1. The resulting design is shown below:



BLOCK	X1	X2
1	1	-0.9
2	1	-1.0
3	1	0.0
4	1	0.9
5	1	0.9
6	1	0.0
7	1	-0.9
8	1	0.0
9	1	1.0
10		
11		
12		
13		
14		
15		
16		
17		
18		
19		
20		
21		
22		

Optimize:
 I-efficiency
 D-efficiency
 A-efficiency
 G-efficiency
 Alias-efficiency

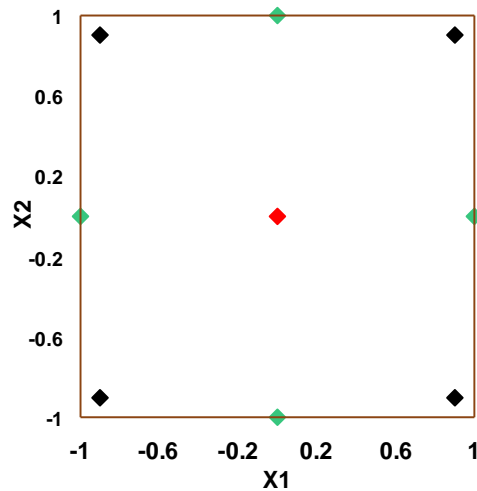
Display:
 Original units
 Coded units
 Randomize run order

Number of coefficients: 6
 Number of base runs: 9
 Number of replicates: 0
 Number of centerpoints: 0
 Group runs in blocks of size: 1000

Average prediction variance: 0.510612
 D-efficiency: 37.54%
 A-efficiency: 23.50%
 trace(A'A): 2.05936

As may be seen in a plot of the design points, the 4 corner points have been pulled in about 10% of the distance from the extreme levels of the factors to the centerpoint:

Response surface design



Note: The generated designs are very sensitive to the selection of λ_D . For example, if λ_D is set to 0.80 instead of 0.81, a much different design is obtained. It is suggested that experimenters try a range of values for λ_D and select from the resulting designs.

References

Box, G. E. P., Hunter, W. G. and Hunter, J. S. (2005). Statistics for Experimenters: An Introduction to Design, Data Analysis, and Model Building, 2nd edition. New York: John Wiley and Sons.

Jones, B. and Nachtsheim, C.J. (2011). "Efficient Designs with Minimal Aliasing." Technometrics 53, pp. 62-71.