Chapter 888

D-Optimal Designs

Introduction

This procedure generates D-optimal designs for multi-factor experiments with both quantitative and qualitative factors. The factors can have a mixed number of levels. For example, you could use this procedure to design an experiment with two quantitative factors having three levels each and a qualitative factor having seven levels.

D-optimal designs are constructed to minimize the generalized variance of the estimated regression coefficients. In the multiple regression setting, the matrix **X** is often used to represent the data matrix of independent variables. D-optimal designs minimize the overall variance of the estimated regression coefficients by maximizing the determinant of **X'X**. Designs that are D-optimal have been shown to be nearly optimal for several other criteria that have been proposed as well.

D-optimal designs are used when there is a limited budget and a completely replicated factorial design cannot be run. For example, suppose you want to study the response to three factors: A with three levels, B with four levels, and C with eight levels. One complete replication of this experiment would require 3 x 4 x 8 = 96 points (we use the word 'point' to mean an experimental unit). If only 20 points can be afforded, the D-optimal design algorithm provides a reasonable choice for the 20 points.

An introduction to experimental design is presented in Chapter 881 on Two-Level Designs and will not be repeated here.

D-Optimal Design Overview

This section provides a brief overview of how the D-optimal design algorithm works. It will provide a general understanding of what the algorithm is trying to accomplish so that you can make intelligent choices for the various options.

Suppose you are studying the influence of height and weight on blood pressure. If you believe that a linear (straight line) relationship exists, you will only need to look at two height values and two weight values. An experiment designed to study this relationship would require four treatment combinations. However, if you decide that the relationship may be curvilinear, you will have to include at least three levels for each factor which results in nine treatment combinations. Clearly, the appropriate experimental design depends on the anticipated functional relationship between the response variable and the factors of interest.

The D-optimal algorithm works as follows. First, specify an approximate mathematical model which defines the functional form of the relationship between the response (Y) and the independent variables (the factors). Next, generate a set of possible candidate points based on this model. Finally, from these candidates select the subset that maximizes the determinant of the **X'X** matrix. This is the D-optimal design. The details of this algorithm are given in Atkinson and Donev (1992).

The number of possible designs grows rapidly as the complexity of the model increases. This number is usually so large that an exhaustive search of all possible designs for a given sample size is not feasible.

The D-optimal algorithm begins with a randomly selected set of points. Points in and out of the current design are exchanged until no exchange can be found that increases the determinant of **X'X**. To cut down on the running time, the number of points considered during any one iteration may be limited.

Unfortunately, this method does not guarantee that the global maximum is found. To overcome this, the algorithm is repeated several times in hopes that at least one iteration leads to the global maximum. For this reason, 50 or 100 random starting sets are needed. (During the testing of the algorithm, we found that some designs required 500 starts to obtain the global maximum.)

Factor Scaling

This algorithm deals with both *quantitative* (continuous) and *qualitative* (discrete) factors. The levels of quantitative factors are scaled so that the minimum value is -1 and the maximum value is 1. Qualitative factors are included as a set of variables. For example, suppose that a qualitative variable has four values. Three independent variables are created to represent this factor:

Original	**X1**	**X2**	**X3**
1	-10	0	0
2	0	-1	0
3	0	0	-1
4	1	1	1

As you can see, each of these variables compares a separate group with the last group. Also note that the number of generated variables is always one less than the number of levels.

Duplicates (Replicates)

The measurement of experimental error is extremely important in the analysis of an experiment. In most cases, if an estimate of experimental error is not available, the data from the experiment cannot be analyzed. One of the best estimates of experimental error comes from points that are duplicates (often called replicates) of each other. Since D-optimal designs are often used in situations with limited budgets, the experimenter is often tempted to ignore the need for duplicates and instead add points with additional treatment combinations. The tenth commandment for experimental design should be "Thou shalt have at least four duplicates in an experiment."

Unfortunately, the D-optimal design algorithm ignores the need for duplicates. Instead, you have to add them after the experimental design has been found. You should set aside at least four points from the algorithm. For example, suppose you have budget for 20 design points. You would tell the program that you have only 16 points. The algorithm would find the best 16-point design. You would then duplicate four of the resulting design points to provide an estimate of experimental error. We recommend that you spread these duplicates out across the experiment so you can have some indication as to whether the magnitude of the experimental error is constant across all treatment settings.

Specifying a Model

Selecting an appropriate model is subjective by nature. Often, you will know very little about the true functional form of the relationship between the response and the factor variables. A common approach is to assume that a second-order Taylor-series approximation will work fairly well. You are assuming that the true function may be approximated by parabolic surface in the neighborhood of interest. Cutting down on the complexity of the model reduces the number of points that must be added to the experimental design.

When dealing with qualitative factors, you generally limit the model to first order interactions. Higher order interactions may be studied later when a complete experiment can be run.

Augmenting an Existing Design

Occasionally, you will want to add more points to an existing experimental design. This may be accomplished by forcing the algorithm to include points that are read from the spreadsheet. The D-optimal algorithm will pick the most useful additional points from the list of candidate points. One of the attractive features of the D-optimal design algorithm is that you can refine the model as your knowledge of it increases.

Example 1 – D-Optimal Design with 10 Points, 3 Factors

This section presents an example of how to generate a D-optimal design using this program. **CAUTION: since the purpose of this routine is to generate data, you should begin with an empty output spreadsheet.**

In this example, we will show you how to generate a 10-point design for a study involving three quantitative factors. We want the design optimized to estimate a second-order response surface model.

Setup

If the procedure window is not already open, use the PASS Home window to open it. The parameters for this example are listed below and are stored in the **Example 1** settings file. To load these settings to the procedure window, click **Open Example Settings File** in the Help Center or File menu.

Design Tab

N Per Block	**10**
Input Columns (Candidate and Forced)	**\<Empty\>**
Number Duplicates	**0**
Input Data Type	**Factor Values**
Forced Points	**None**
Optimize the Design for this Model	**A|A|B|B|C|C**
Max Term Order	**2**
Qualitative Factors and Levels	**\<Empty\>**
Max Iterations	**30**
Inclusion Points	**5**
Removal Points	**5**
Random Seed	**2119855**
Include Intercept	**Checked**
Store Data on Spreadsheet	**Checked**
First Factor Column	**1**
First Expanded Column	**5**
Rename Factor Columns with Factor Labels	**Checked**
Clear Existing Data	**Checked**

Output

Click the Calculate button to perform the calculations and generate the following output.

10-Point, 3 Factor D-Optimal Design on Spreadsheet

A	B	C	C4	Int't	Ax	Bx	CxA_A	A_B	A_C	B_B	B_C	C_C
-1	-1	-1	1	-1	-1	-1	1	1	1	1	1	1
1	-1	-1	1	1	-1	-1	1	-1	-1	1	1	1
0	0	-1	1	0	0	-1	0	0	0	0	0	1
-1	1	-1	1	-1	1	-1	1	-1	1	1	-1	1
1	1	-1	1	1	1	-1	1	1	-1	1	-1	1
1	0	0	1	1	0	0	1	0	0	0	0	0
0	1	0	1	0	1	0	0	0	0	1	0	0
1	-1	1	1	1	-1	1	1	-1	1	1	-1	1
-1	0	1	1	-1	0	1	1	0	-1	0	0	1
1	1	1	1	1	1	1	1	1	1	1	1	1

Several columns in the spreadsheet are filled with data. The results may vary from the values seen here, as there may be equally optimal designs. The first, second, and third, variables (A, B, and C) contain the actual design. You would replace the -1's with the corresponding factor's minimum value, the 1's with the maximum value, and the 0's with the average of the two.

The variables from Intercept to C_C contain the expanded design matrix. Each variable is generated by multiplying the appropriate factor values. For example, in the first row, A_B is found by multiplying the value for A, which is -1, by the value for B, which is also -1. The result is 1. The intercept is set to one for all rows. The expanded matrix is usually saved so that the design can be analyzed using multiple regression.

To use this design, you would randomly assign these ten points to the ten experimental units.

Factor Section

Factor Section

Name	Number Values	Type	Value1	Value2	Value3
A	3	Quantitative	-1	0	1
B	3	Quantitative	-1	0	1
C	3	Quantitative	-1	0	1

A total of 27 observations will be needed for one replication.

This report summarizes the factors that were included in the design. The last line of this report gives the number of observations required for one complete replication of the experiment. This value is the product of the number of levels for each factor.

Name

The symbol(s) used to represent the factor.

Number Values

The number of values (levels) generated for each factor. For qualitative factors, this value was set in the Qualitative Factors and Levels box of the Design panel. For quantitative factors, this value is one more that the highest exponent used with this term. For example, if the model includes an A*A and nothing of a higher order, this value will be three.

Type

A factor is either quantitative or qualitative.

Value1 – Value 3

These columns list the individual values that are used as the levels of each factor when generating the expanded design matrix based on the model. Notice that the smallest is always -1 and the largest is always 1.

When the expanded design matrix is input directly, these values should be ignored.

Model Terms Section

Model Terms Section

Variables Needed	Term
1	A
1	B
1	C
1	A*A
1	A*B
1	A*C
1	B*B
1	B*C
1	C*C
9	Model Total

This report shows the terms generated by your model. You should check this report carefully to make sure that the generated model matches what you wanted. The last line of the report gives the total number of degrees of freedom (except for the intercept) required for your model. This number plus one is the minimum size of the D-optimal design for this model.

Variables Needed

The number of degrees of freedom (expanded design variables) required for this term.

Term

The name of each term.

D-Optimal Design

D-Optimal Design

Original Row	Factor A	Factor B	Factor C
1	-1	-1	-1
3	1	-1	-1
5	0	0	-1
7	-1	1	-1
9	1	1	-1
13	-1	0	0
17	0	1	0
19	-1	-1	1
24	1	0	1
25	-1	1	1

Computer-Generated Random Seed: 2119855

This report gives the points in the D-optimal design.

Original Row

This is the row number of the point from the list of candidate points. It is only useful in those cases in which you provided the list of candidate points manually.

Factors (A B C)

These are the values of the factors. For example, the first-row sets A, B, and C to -1. Remember that these are scaled values. You would transform them back into their original metric using the formula:

Original = (Scaled(Max - Min) + Max + Min)/2

For example, suppose the original metric for factor A is minimum = 10 and maximum =20. The original values would be calculated as follows:

Scaled	Formula	Original
-1	(-1(20-10)+20+10)/2	10
0	(0(20-10)+20+10)/2	15
1	(1(20-10)+20+10)/2	20

The values 10, 15, and 20 represent the three levels of factor A that are used in the design. They would replace the -1, 0, and 1 displayed in this report.

Determinant Analysis Section

Determinant Analysis Section

Rank	Determinant of X'X	D-Efficiency	Percent of Maximum
1	1327104	40.95	100.00
2	1327104	40.95	100.00
3	1327104	40.95	100.00
4	1327104	40.95	100.00
5	1327104	40.95	100.00
6	1048576	40.00	79.01
7	1048576	40.00	79.01
8	1048576	40.00	79.01
9	1048576	40.00	79.01
10	1048576	40.00	79.01
11	1048576	40.00	79.01
12	1048576	40.00	79.01
13	921600	39.49	69.44
14	921600	39.49	69.44
15	921600	39.49	69.44
16	802816	38.95	60.49
17	802816	38.95	60.49
18	802816	38.95	60.49
19	802816	38.95	60.49
20	802816	38.95	60.49

The maximum was achieved on 5 of 30 iterations.
Computer-Generated Random Seed: 2119855

This report shows the largest twenty determinants. The main purpose of this report is to let you decide if enough iterations have been run so that a global maximum has been found. Unless the maximum value was achieved on at least five iterations, you should double the number of iterations and rerun the procedure.

In this example, the top value occurred on only two iterations. In practice we would probably try another 200 iterations to find out if this is the global maximum.

Rank

Only the top twenty are shown on this report. The values are sorted by the determinant.

Determinant of X'X

This is the value of the determinant of **X'X** which is the statistic that is being maximized. This value is sometimes called the generalized variance of the regression coefficients. Since this value occurs in the denominator of the variance of each regression coefficient, maximizing it has the effect of reducing the variance of the estimated regression coefficients.

D-Efficiency

D-efficiency is the relative number of runs (expressed as a percent) required by a hypothetical orthogonal design to achieve the same determinant value. It provides a way of comparing designs across different sample sizes.

$$DE = 100 \left(\frac{|X'X|^{1/p}}{N} \right)$$

where p is the total number of degrees of freedom in the model and N is the number of points in the design.

Percent of Maximum

This is the percentage that the determinant on this row is of the best determinant found.

Individual Degree of Freedom Section

Individual Degree of Freedom Section

Number	Name	Diagonal of X'X	Diagonal of X'X Inv
1	Intercept	10	0.861111
2	A	8	0.166667
3	B	7	0.250000
4	C	8	0.166667
5	A*A	8	0.861111
6	A*B	6	0.250000
7	A*C	7	0.194444
8	B*B	7	0.722222
9	B*C	6	0.250000
10	C*C	8	0.861111

Determinant	1327104
D-Efficiency	40.95345
Trace	4.583333
A-Efficiency	21.81818

This report shows the diagonal elements of the **X'X** and its inverse. Since the variance of each term is proportional to diagonal elements from the inverse of **X'X**, the last column of this report lets you compare those variances. From this report you can determine if the coefficients will be estimated with the relative precision that is desired.

For example, we can see from this example that them main effects will be estimated with the greatest precision—usually a desirable quality in a design.

Number

An arbitrary sequence number.

Name

The name of the term.

Diagonal of X'X

The diagonal element of this term in the **X'X** matrix.

Diagonal of X'X Inv

The diagonal element of this term in the **X'X** inverse matrix. See the discussion above for an understanding of how this value might be interpreted.

Determinant

This is the value of the determinant of **X'X** which is the statistic that is being maximized. This value is sometimes called the generalized variance of the regression coefficients. Since this value occurs in the denominator of the variance of each regression coefficient, maximizing it has the effect of reducing the variance of the estimated regression coefficients.

D-Efficiency

D-efficiency is the relative number of runs (expressed as a percent) required by a hypothetical orthogonal design to achieve the same determinant value. It provides a way of comparing designs across different sample sizes.

$$DE = 100 \left(\frac{|X'X|^{1/p}}{N} \right)$$

where p is the total number of degrees of freedom in the model and N is the number of points in the design.

Trace

This is the value of the trace of **X'X-inverse** which is associated with A-optimality.

A-Efficiency

D-efficiency is the relative number of runs (expressed as a percent) required by a hypothetical orthogonal design to achieve the same trace value. It provides a way of comparing designs across different sample sizes.

$$AE = 100 \left(\frac{p}{trace(N(X'X)^{-1})} \right)$$

where p is the total number of degrees of freedom in the model and N is the number of points in the design.

Candidate Points Section

Candidate Points Section

Original Row	Factor A	Factor B	Factor C
1	-1	-1	-1
2	0	-1	-1
3	1	-1	-1
4	-1	0	-1
5	0	0	-1
6	1	0	-1
7	-1	1	-1
8	0	1	-1
9	1	1	-1
10	-1	-1	0
11	0	-1	0
12	1	-1	0
13	-1	0	0
14	0	0	0
15	1	0	0
16	-1	1	0
17	0	1	0
18	1	1	0
19	-1	-1	1
20	0	-1	1
21	1	-1	1
22	-1	0	1
23	0	0	1
24	1	0	1
25	-1	1	1
26	0	1	1
27	1	1	1

This report gives a list of candidate points from which the D-optimal design points were selected.

Original Row

This is an arbitrary identification number.

Factors (A B C)

These are the values of the factors. For example, the first-row sets A, B, and C to -1. Remember that these are scaled values. You would transform them back into their original metric using the formula:

$$\text{Original} = (\text{Scaled}(\text{Max} - \text{Min}) + \text{Max} + \text{Min})/2$$

For example, suppose the original metric for factor A is minimum = 10 and maximum =20. The original values would be calculated as follows:

Scaled	Formula	Original
-1	(-1(20-10)+20+10)/2	10
0	(0(20-10)+20+10)/2	15
1	(1(20-10)+20+10)/2	20

The values 10, 15, and 20 represent the three levels of factor A. They would replace the -1, 0, and 1 displayed in this report.

Expanded Design Matrix Section

Expanded Design Matrix Section

Row	Intercept	A	B	C	A*A	A*B	A*C	B*B	B*C	C*C
1	1	-1	-1	-1	1	1	1	1	1	1
2	1	0	-1	-1	0	0	0	1	1	1
3	1	1	-1	-1	1	-1	-1	1	1	1
4	1	-1	0	-1	1	0	1	0	0	1
5	1	0	0	-1	0	0	0	0	0	1
6	1	1	0	-1	1	0	-1	0	0	1
7	1	-1	1	-1	1	-1	1	1	-1	1
8	1	0	1	-1	0	0	0	1	-1	1
9	1	1	1	-1	1	1	-1	1	-1	1
10	1	-1	-1	0	1	1	0	1	0	0
11	1	0	-1	0	0	0	0	1	0	0
12	1	1	-1	0	1	-1	0	1	0	0
13	1	-1	0	0	1	0	0	0	0	0
14	1	0	0	0	0	0	0	0	0	0
15	1	1	0	0	1	0	0	0	0	0
16	1	-1	1	0	1	-1	0	1	0	0
17	1	0	1	0	0	0	0	1	0	0
18	1	1	1	0	1	1	0	1	0	0
19	1	-1	-1	1	1	1	-1	1	-1	1
20	1	0	-1	1	0	0	0	1	-1	1
21	1	1	-1	1	1	-1	1	1	-1	1
22	1	-1	0	1	1	0	-1	0	0	1
23	1	0	0	1	0	0	0	0	0	1
24	1	1	0	1	1	0	1	0	0	1
25	1	-1	1	1	1	-1	-1	1	1	1
26	1	0	1	1	0	0	0	1	1	1
27	1	1	1	1	1	1	1	1	1	1

This report gives a list of candidate points expanded so that each individual term may be seen. The report is useful to show you how the expanded matrix looks. Each variable is generated by multiplying the appropriate factor values. For example, in the first row, A_B is found by multiplying the value for A, which is -1, by the value for B, which is also -1. The result is 1. The intercept is set to one for all rows.

If you want to constrain the design space, you could cut and paste these values back into the spreadsheet and then eliminate points that cannot occur.

Example 2 – Two Factors

This section presents an example of how to generate and analyze a D-optimal design involving two factors. Suppose we want to study the effect of two factor variables, A and B, on a response variable, Y. A and B happen to be quantitative variables and there is reason to believe that a second-order response surface design will work well. A full replication of this design requires nine points. In addition, four more are required to provide an estimate of experimental error. However, we can only afford eight. We will create a D-optimal design with six of the experimental units and use the remaining two as duplicates to provide the estimate of experimental error.

We want to analyze the response surface for values of A between 10 and 20 and values of B between 1 and 3.

Setup

If the procedure window is not already open, use the PASS Home window to open it. The parameters for this example are listed below and are stored in the **Example 2** settings file. To load these settings to the procedure window, click **Open Example Settings File** in the Help Center or File menu.

Design Tab

N Per Block **6**
Input Columns (Candidate and Forced) .... **<Empty>**
Number Duplicates **0**
Input Data Type .... **Factor Values**
Forced Points .... **None**
Optimize the Design for this Model .... **A|A|B|B**
Max Term Order **2**
Qualitative Factors and Levels .... **<Empty>**
Max Iterations **30**
Inclusion Points **5**
Removal Points **5**
Random Seed .... **2218619** (for Reproducibility)
Include Intercept .... **Checked**
Store Data on Spreadsheet .... **Checked**
First Factor Column **1**
First Expanded Column **4**
Rename Factor Columns with Factor Labels .... **Checked**
Clear Existing Data .... **Checked**

Output

Click the Calculate button to perform the calculations and generate the following output.

6-Point, 2 Factor D-Optimal Design

D-Optimal Design

Original Row	Factor A	Factor B
1	-1	-1
3	1	-1
5	0	0
6	1	0
7	-1	1
9	1	1

Computer-Generated Random Seed: 2218619

Columns A and B give the design. The Determinant Analysis Section showed that the maximum was achieved in less than 30 iterations. Hence, we assume that the algorithm converged to the global maximum.

Next, we manually add the two duplicates to the design. When only a few duplicates are available, we like to have them in the middle, so we will duplicate the two rows having zero values. We choose random numbers for the two new response values. The resulting design appears as follows.

6-Point Design with Two Duplicates

D-Optimal Design

Original Row	Factor A	Factor B
1	-1	-1
3	1	-1
5	0	0
5	0	0
6	1	0
6	1	0
7	-1	1
9	1	1

Computer-Generated Random Seed: 2218619

Next, we change the factor values back to their original scale. Factor A went from 10 to 20 and factor B went from 1 to 3. The resulting database appears as follows.

6-Point Design in Original Scale

D-Optimal Design

Original Row	Factor A1	Factor B1
1	10	1
3	20	1
5	15	2
5	15	2
6	20	2
6	20	2
7	10	3
9	20	3

Computer-Generated Random Seed: 2218619

We could continue this exercise by running these data through a multiple regression program, paying particular attention to the Multicollinearity Section and the Eigenvalues of Centered Correlations Section. When we did this, we found that multicollinearity seemed to be a problem in the original scale, but not in the -1 to 1 scale used by the D-optimal algorithm.

Example 3 – Three Factors with Blocking

This section presents an example of how to generate and analyze a D-optimal design involving three factors with blocking.

Suppose we want to study the effect of three quantitative factor variables (A, B, and C) on a response variable. There is reason to believe that a second-order response surface design will work well. A full replication of this design requires twenty-seven experimental units. The manufacturing process that we are studying produces items in batches of four at a time. Because of this and the limited budget available for this study, we decide to use three batches (which we will call 'Blocks') of four points each.

Setup

If the procedure window is not already open, use the PASS Home window to open it. The parameters for this example are listed below and are stored in the **Example 3** settings file. To load these settings to the procedure window, click **Open Example Settings File** in the Help Center or File menu.

Design Tab

N Per Block..**4 4 4**
Input Columns (Candidate and Forced)................**<Empty>**
Number Duplicates **0**
Input Data Type ....**Factor Values**
Forced Points..**None**
Optimize the Design for this Model ....**A|B|C A*A B*B C*C**
Max Term Order **2**
Qualitative Factors and Levels.............................**<Empty>**
Max Iterations **100**
Inclusion Points..**45**
Removal Points..**11**
Random Seed...**2400459** (for Reproducibility)
Include Intercept ....**Checked**
Store Data on Spreadsheet ....**Checked**
First Factor Column **1**
First Expanded Column ....**<Empty>**
Rename Factor Columns with Factor Labels........**Checked**
Clear Existing Data ....**Checked**

Output

Click the Calculate button to perform the calculations and generate the following output.

12-Point, 3 Factor D-Optimal Design with Blocking

D-Optimal Design

Original Row	Factor A	Factor B	Factor C	Block
2	-1	-1	-1	2
9	1	-1	-1	3
14	0	0	-1	2
21	-1	1	-1	3
25	1	1	-1	1
37	-1	0	0	1
49	0	1	0	1
53	1	1	0	2
57	-1	-1	1	3
61	1	-1	1	1
74	-1	1	1	2
81	1	1	1	3

User-Entered Random Seed: 2400459

Columns A, B, C, and Blocks give the design. The Determinant Analysis Section showed that the maximum was achieved on 10 of the 100 iterations. Hence, we assume that the algorithm converged to the global maximum.

Example 4 – Adding Points to an Existing Design

This section presents an example of how to augment additional points to an existing design.

Suppose a standard three factor design has been run. Each factor has two levels. The design was blocked into two blocks of four points each. The design values are contained in the spreadsheet when the example is loaded. This design allows only first-order (linear) terms to be fit.

Suppose that you wish to add more points to the design so that a second-order response surface may be fit. Specifically, suppose you want to add one more block of four points to extend the model from first to second order. What four points should be added?

Setup

If the procedure window is not already open, use the PASS Home window to open it. The parameters for this example are listed below and are stored in the **Example 4** settings file. To load these settings to the procedure window, click **Open Example Settings File** in the Help Center or File menu.

Design Tab

N Per Block..**4 4 4**
Input Columns (Candidate and Forced)..................**A-C**
Number Duplicates **0**
Input Data Type ....**Factor Values**
Forced Points..**8**
Optimize the Design for this Model ....**A|B|C A*A B*B C*C**
Max Term Order **2**
Qualitative Factors and Levels..............................**<Empty>**
Max Iterations **30**
Inclusion Points...**5**
Removal Points..**5**
Random Seed..**2455768** (for Reproducibility)
Include Intercept ....**Checked**
Store Data on Spreadsheet ....**Checked**
First Factor Column **1**
First Expanded Column ....**<Empty>**
Rename Factor Columns with Factor Labels ....**Checked**
Clear Existing Data ....**Checked**

Output

Click the Calculate button to perform the calculations and generate the following output.

Augmented D-Optimal Design with Blocking

D-Optimal Design

Original Row	Factor A	Factor B	Factor C	Block
1	-1	-1	-1	1
2	1	1	-1	1
3	1	-1	1	1
4	-1	1	1	1
5	1	-1	-1	2
6	-1	1	-1	2
7	-1	-1	1	2
8	1	1	1	2
12	-1	0	-1	3
16	0	1	-1	3
18	-1	-1	0	3
22	0	0	0	3

User-Entered Random Seed: 2455768

Variables A, B, C, and Blocks give the design. The new block is shown as the last four rows of the design.

The Determinant Analysis Section showed that the maximum was achieved on 9 of the 30 iterations. Hence, we assume that the algorithm converged to the global maximum.

Example 5 – Mixture Design

This section presents an example of how to generate a mixture design. Mixture designs are useful in situations in which the factors are constrained to sum to a total. The interest is in the proportions of each factor, not the absolute amounts. For example, the proportions of the components of a chemical solution must sum to one.

Suppose that you wish to design a first-order mixture experiment for a chemical that has three components (which we will label as A, B, and C). In this case, you will not code the factor levels from -1 to 1. Rather, the factor levels will be coded from zero to one. Because of this constraint, the intercept will not be fit in this model.

In this particular case, we will constrain the design space by only entering certain points in the list of candidate points. The candidate points are contained in the input spreadsheet when the example is loaded. The following plots (not generated using *PASS*) show the design space for each pair of factors. Remember that these factors are constrained so that the missing factor is equal to one minus the sum of the other two. Hence, if A is 0.7 and B is 0.2, then C must be 0.1.
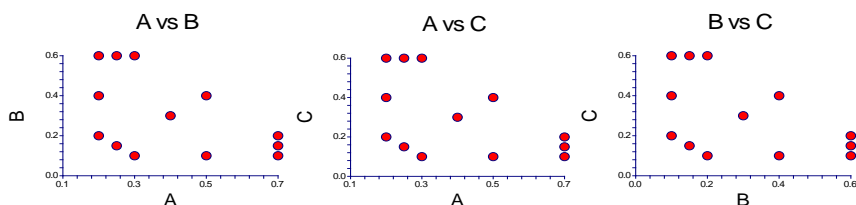
The task for the algorithm is to pick the ten best points from the thirteen that are shown here.

Setup

If the procedure window is not already open, use the PASS Home window to open it. The parameters for this example are listed below and are stored in the **Example 5** settings file. To load these settings to the procedure window, click **Open Example Settings File** in the Help Center or File menu.

Design Tab

N Per Block..**10**
Input Columns (Candidate and Forced)................**1-3**
Number Duplicates **0**
Input Data Type ....**Factor Values**
Forced Points...**0**
Optimize the Design for this Model ....**A|B|C**
Max Term Order **2**
Qualitative Factors and Levels.............................**<Empty>**
Max Iterations **30**
Inclusion Points..**5**
Removal Points...**5**
Random Seed..**2498700** (for Reproducibility)
Include Intercept ....**Unchecked**
Store Data on Spreadsheet ....**Checked**

First Factor Column **1**
First Expanded Column ....**<Empty>**
Rename Factor Columns with Factor Labels........**Checked**
Clear Existing Data ....**Checked**
──

Input Spreadsheet Data
──

Row	A	B	C
1	0.70	0.10	0.20
2	0.20	0.60	0.20
3	0.70	0.20	0.10
4	0.20	0.20	0.60
5	0.30	0.60	0.10
6	0.30	0.10	0.60
7	0.70	0.15	0.15
8	0.20	0.40	0.40
9	0.50	0.10	0.40
10	0.25	0.60	0.15
11	0.50	0.40	0.10
12	0.25	0.15	0.60
13	0.40	0.30	0.30
──

Output

Click the Calculate button to perform the calculations and generate the following output.

Mixture Design

D-Optimal Design
──

Original Row	Factor A	Factor B	Factor C
1	0.7	0.1	0.2
2	0.2	0.6	0.2
3	0.7	0.2	0.1
4	0.2	0.2	0.6
5	0.3	0.6	0.1
6	0.3	0.1	0.6
8	0.2	0.4	0.4
9	0.5	0.1	0.4
11	0.5	0.4	0.1
13	0.4	0.3	0.3
──

User-Entered Random Seed: 2498700

Columns A, B, and C give the design. The original row from the candidate list is shown as the first column of the report.

The Determinant Analysis Section showed that the maximum was achieved on 30 of the 30 iterations. Hence, we assume that the algorithm did not converge to the global maximum.

In order to visually analyze the design, we generate the scatter plots (not generated using *PASS*) for each pair of variables in the design.

Plot of Design
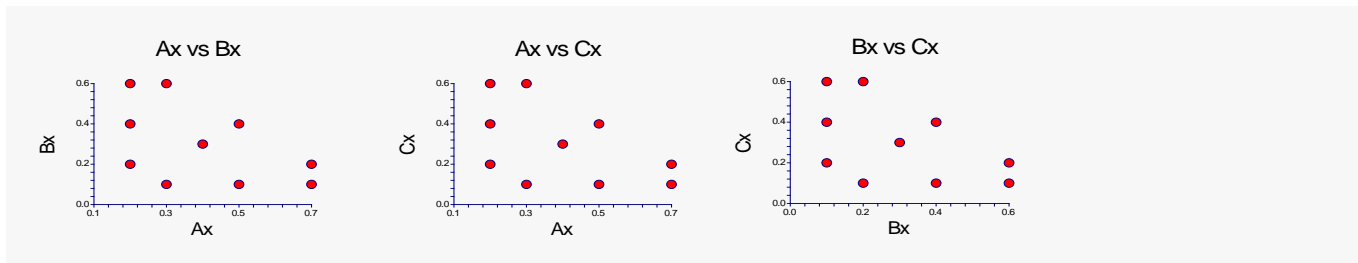
It is interesting to compare these plots with those produced earlier to see which points were kept by the algorithm.

Example 6 – Qualitative Factors

This section presents an example of how to design an experiment with qualitative and quantitative factors.

Suppose your experimental situation involves two quantitative variables, A and B, and a qualitative variable C that has five possible levels. You want to fit a second-order response surface to the quantitative variables. Also, you want to fit all two-way interactions among these factors. You have budget for an 18-point design (you will add four duplicates later).

Setup

If the procedure window is not already open, use the PASS Home window to open it. The parameters for this example are listed below and are stored in the **Example 6** settings file. To load these settings to the procedure window, click **Open Example Settings File** in the Help Center or File menu.

Design Tab

N Per Block...	**18**
Input Columns (Candidate and Forced)..................	**<Empty>**
Number Duplicates	**0**
Input Data Type	**Factor Values**
Forced Points...	**0**
Optimize the Design for this Model	**A|B|C A*A B*B**
Max Term Order	**2**
Qualitative Factors and Levels................................	**C(5)**
Max Iterations	**30**
Inclusion Points..	**20**
Removal Points..	**18**
Random Seed...	**2498700** (for Reproducibility)
Include Intercept	**Unchecked**
Store Data on Spreadsheet	**Checked**
First Factor Column	**1**
First Expanded Column	**<Empty>**
Rename Factor Columns with Factor Labels	**Checked**
Clear Existing Data	**Checked**

Output

Click the Calculate button to perform the calculations and generate the following output.

Design with Qualitative Factors

D-Optimal Design

Original Row	Factor A	Factor B	Factor C
2	0	-1	1
4	-1	0	1
6	1	0	1
8	0	1	1
10	-1	-1	2
12	1	-1	2
16	-1	1	2
18	1	1	2
19	-1	-1	3
21	1	-1	3
25	-1	1	3
28	-1	-1	4
30	1	-1	4
36	1	1	4
37	-1	-1	5
39	1	-1	5
43	-1	1	5
45	1	1	5

User-Entered Random Seed: 2498700

Columns A, B, and C give the design. Notice that column C simply gives the level for factor C—it was not rescaled. Also note that the levels of factor C are numbered arbitrarily. This means that only the pattern is important, not the particular level. For example, in this solution, there are only three level 2's and three level 4's. In the next solution, there might be three level 3's and three level 4's.

The Determinant Analysis Section showed that the maximum was achieved on 5 of the 30 iterations. Hence, we assume that the algorithm converged to the global maximum.