

## 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  $\mathbf{X}$  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  $\mathbf{X}'\mathbf{X}$ . 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  $3 \times 4 \times 8 = 96$  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 Factorial 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  $\mathbf{X}'\mathbf{X}$  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  $\mathbf{X}'\mathbf{X}$ . To cut down on the running time, the number of points considered during any one iteration may be limited.

## D-Optimal Designs

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:

<u>Original</u>	<u>X1</u>	<u>X2</u>	<u>X3</u>
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

## D-Optimal Designs

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.

## Procedure Options

This section describes the options available in this procedure.

### Design Tab

This panel specifies the parameters that will be used to create the design values.

### Experimental Setup

#### N Per Block

This option specifies the required sample size. If you are not using blocks, enter a single number giving the total sample size. The sample size must be large enough to fit the designated model. If it is not large enough, you will be shown the minimum number of points necessary.

If you are using blocks, enter the sample size for each block, separated by blanks or commas. These sample sizes do not have to be equal, although they usually are. For example, if you have three blocks, you might enter 8,8,12 which would give an overall sample size of 28. The first block will have 8 points, the second 8 points, and the third 12 points.

You must be careful when specifying blocks when you also have forced design points. In this case, the first few blocks are matched with the forced design points. The size of the blocks must match the number of forced points. For example, suppose you have already run two blocks of four each and you want to augment this with three blocks of six each. You would have eight forced points. The entry in this field would be 4,4,6,6,6. If you entered 4,3,7,6,6 an error would occur because the forced points cannot be assigned exactly to one or more blocks. The bottom line is, you cannot force partial blocks into the design.

#### Input Columns (Candidate and Forced)

When specified, these columns contain either a set of points to be forced into the final design, a set of candidate points from which the design is to be selected, or both. The data must be arranged so that the forced points are located at the top of the spreadsheet followed by any candidate points. When candidate points are specified, no additional candidate points are generated. If you want to force points in the design and choose the rest from among those generated by the model statement, the total number of rows in these variables must equal the total number of forced rows specified below.

Note that these variables are matched with the factors specified in the model after those factors have been sorted.

Qualitative factors must be entered using positive integers (1, 2, 3, etc.). You cannot use any other identifiers. If you have data entered using some other scheme (such as A, B, C, etc.), you will have to recode the values so that they are positive integers.

Quantitative factors must be scaled so that the minimum value is -1 and the maximum value is 1. For example, suppose an existing design has a factor whose values are 10, 15, and 20. Here the minimum is 10 and the maximum is 20. You would transform these using the formula

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

Since, in this example, Max = 20 and Min = 10, the transformation reduces to  $\text{New} = (\text{Original} - 10) / 10 = \text{Original} / 10 - 1$ . You would create a new variable using the transformation  $\text{Original} / 10 - 1$ . This transformation would give  $10/10 - 1 = -1$ ,  $15/10 - 1 = 0.5$ , and  $20/10 - 1 = 1$ . That is, the new variable would contain -1's, 0's, and 1's instead of 10's, 15's, and 20's.

## D-Optimal Designs

### Number Duplicates

It is very important to have duplicates of at least some of the design points to provide an estimate of experimental error. This option designates the number of duplicates to be generated. The first design point is duplicated, then the second, and so on. Even though this option is convenient, we recommend that you pick appropriate points for duplication by looking at scatter plots of the design.

If your design includes blocking, you should not create duplicates since that will give erroneous block sizes. Rather, you should manually create duplicates.

### Input Data Type

If you have Input Variables specified, this option specifies the type of data contained in those variables. Two types of data are possible.

- **Factor Values**

Specifies that the input data contains indices of each factor. An expanded design matrix will be generated from these factor indices using the designated model. This is the more common data type.

- **Expanded Matrix**

Specifies that the input dataset contains the expanded design matrix. That is, the quadratic, cubic, and interaction terms have been created. The model statement is not used. You would use this option when you want to specify the candidate design set in more detail than is allowed by the program. The expanded matrix must include the intercept (a column of one's) if one is to be included in the model.

### Forced Points

The number of rows in the Input Variables that should be forced into the final design. These rows must be located at the top of the database, before any candidate points. If the number of forced points is equal to the number of points read in, the generated design matrix is used. Otherwise, the additional rows are used as candidate points and no other rows are generated.

### Optimize the Design for this Model

Your design is optimized for the model specified here. Specify main effects (factors) with names consisting of one or more letters, such as A B C. Specify interactions using an asterisk (\*), such as A\*B. You can use the bar (|) symbol (see examples below) as a shorthand method to specify a complete model. You can use parentheses. You can separate terms with blanks or the '+' (plus) sign. Duplicate terms are removed during the evaluation of the model. Note that the main effects are always sorted in alphabetical order.

Some examples will help to indicate how the model syntax works:

$C + B + A + B*A + C*A$	=	$A+B+C+A*B+A*C$ (Note the sorting!)
$A B$	=	$A+B+A*B$
$B A$	=	$A+B+A*B$
$A B A*A B*B$	=	$A+B+A*B+A*A+B*B$
$A A B B$ (Max Term Order=2)	=	$A+B+A*B+A*A+B*B$
$A B C$	=	$A+B+C+A*B+A*C+B*C+A*B*C$
$(A+B)*(C+D)$	=	$A*C+A*D+B*C+B*D$
$(A+B) C$	=	$A+B+C+(A+B)*C$
	=	$A+B+C+A*C+B*C$

You can experiment with various expressions by viewing the Model Terms report.

For quantitative factors, each term represents a single variable in the expanded design matrix. For qualitative variables, each term represents a set of variables in the expanded design matrix.

## D-Optimal Designs

Note that qualitative terms should not be squared or cubed. That is, if A is a qualitative factor, you would not include  $A^2$  or an  $A^3$  in your model.

### Max Term Order

This option specifies that maximum number of factors that can occur in an interaction term. For example,  $A^2B^2C$  is a third order interaction term and if this option were set to 2, the  $A^2B^2C$  would be removed from the model.

This option is particularly useful when used with the bar notation to remove unwanted terms.

### Qualitative Factors and Levels

List any qualitative factors here followed by the number of levels given in parenthesis. Factors in the model which do not appear here are assumed to be quantitative (continuous). For example, you might enter  $A(5), B(4), C(7)$  to indicate three qualitative factors, one with five levels, the next with four levels, and the third with seven levels. Of course, the names used here must match the names used in the model statement.

### Max Iterations

Specify the number of times the algorithm is started with a new random design. Often 50 or 100 iterations are necessary and 500 is not unheard of. As the number of Inclusion Points and Removal Points are increased (see below), the number of iterations may be decreased.

We suggest that you increase this value until the optimal design is found on several iterations as reported in the Determinant Analysis report.

### Inclusion Points

This is the number of candidate points considered for addition during an iteration. Instead of considering all candidate points, only this many are used. A value between 1 and  $N_c - 1$  (where  $N_c$  is the number of candidate points) may be used. Usually, a value near  $N_c/2$  is adequate.

### Removal Points

This is the number of points currently in the design that will be considered for removal during a particular iteration. A value between 1 and  $N$  (the desired sample size) is used. Setting this value smaller than  $N$  speeds up the search, but reduces the possibility of finding the optimal design.

### Include Intercept

This option specifies whether to include the intercept in the expanded design matrix. Usually, the intercept is left out of mixture designs. The intercept is automatically deleted in designs with more than one block.

---

## Data Storage to Spreadsheet

### Store Data on Spreadsheet

Check this box to generate the design data on the spreadsheet. The spreadsheet data will be identical to the design data generated on the output window.

### First Factor Column

If the Input Data Type is set to Factor Values, the final design is stored in a set of contiguous columns of the spreadsheet, beginning with this column. Be careful not to overwrite existing data. If you have four factors, the design will be stored in this variable and the next three to the right. Existing data will be lost!

If the Input Data Type is set to Expanded Matrix, an index is stored in this column that represents whether the row is used in the design. If the row is not in the optimum design, a zero is stored. If the row is in the optimum design, the number of times it occurs is stored here.

## D-Optimal Designs

### First Expanded Column

This option specifies the first column in which to store the expanded version of the selected design. The rest of the expanded design columns will be stored in the columns to the right. Use this option if you want to output the expanded design matrix for use in a multiple regression procedure.

Warning: The program fills these columns with data, so existing data will be replaced.

---

## Data Storage to Spreadsheet – Storage Options

### Rename Factor Columns with Factor Labels

The names of the factors that were used in the model statement are used to rename the columns in which the design is stored.

### Clear Existing Data

Clear all existing data in the design columns before writing the new design data. This is especially useful if you are experimenting with several designs of different sizes. You will not be warned that data is being lost. The data will be cleared and the new design written automatically.

---

## 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

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the **D-Optimal Designs** procedure window by expanding **Design of Experiments**, then clicking on **Experimental Design**, and then clicking on **D-Optimal Designs**. You may then make the appropriate entries as listed below, or open **Example 1** by going to the **File** menu and choosing **Open Example Template**.

### 1 Specify the design and data storage.

- On the D-Optimal Designs window, select the **Design tab**.
- Set **N per Block** to **10**.
- Set **Optimize the Design for this Model** to **A|A|B|B|C|C**.
- Set **Max Term Order** to **2**.
- Set the **First Factor Column** to **1**.
- Set the **First Expanded Column** to **5**.
- Check **Rename Factor Column with Factor Labels**.
- Check **Clear Existing Data**.

### 2 Specify the reports.

- On the D-Optimal Designs window, select the **Reports tab**.
- Check all reports.
- Set **Decimal Places** to **0**.

### 3 Run the procedure.

- From the Run menu, select **Start Calculation**. Alternatively, just click the Calculate button.

## D-Optimal Designs

## 10-Point, 3 Factor D-Optimal Design

A	B	C	C4	Int't	Ax	Bx	Cx	A_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	1	1
0	0	-1	1	0	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	1
1	0	0	1	1	0	0	1	0	0	0	0	0	0
0	1	0	1	0	1	0	0	0	0	1	1	0	0
1	-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	0	1
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

Name	Number Values	Type	Value1	Value2	Value3
A	3	Quantitative	-1.0000	0.0000	1.0000
B	3	Quantitative	-1.0000	0.0000	1.0000
C	3	Quantitative	-1.0000	0.0000	1.0000

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 than 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

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

Original Row	Factors		
	A	B	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
20	0	-1	1
25	-1	1	1
27	1	1	1

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:

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

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



## D-Optimal Designs

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

Rank	Determinant of X'X	D-Efficiency	Percent of Maximum
1	1327104	40.95	100.00
2	1327104	40.95	100.00
3	1048576	40.00	79.01
4	1048576	40.00	79.01
5	1048576	40.00	79.01
6	1048576	40.00	79.01
7	1048576	40.00	79.01
8	921600	39.49	69.44
9	921600	39.49	69.44
10	802816	38.95	60.49
11	802816	38.95	60.49
12	802816	38.95	60.49
13	802816	38.95	60.49
14	802816	38.95	60.49
15	802816	38.95	60.49
16	746496	38.66	56.25
17	589824	37.76	44.44
18	589824	37.76	44.44
19	589824	37.76	44.44
20	589824	37.76	44.44

The maximum was achieved on 2 of 30 iterations.

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

Number	Name	Diagonal of X'X	Diagonal of X'X Inv
1	Intercept	10.0000	0.861111
2	A	7.0000	0.250000
3	B	8.0000	0.166667
4	C	8.0000	0.166667
5	A*A	7.0000	0.722222
6	A*B	6.0000	0.250000
7	A*C	6.0000	0.250000
8	B*B	8.0000	0.861111
9	B*C	7.0000	0.194444
10	C*C	8.0000	0.861111
Determinant		1327104	
D-Efficiency		40.95345	
Trace		4.583333	
A-Efficiency		21.81818	

This report shows the diagonal elements of the  $\mathbf{X}'\mathbf{X}$  and its inverse. Since the variance of each term is proportional to diagonal elements from the inverse of  $\mathbf{X}'\mathbf{X}$ , 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 the 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  $\mathbf{X}'\mathbf{X}$  matrix.

### Diagonal of X'X Inv

The diagonal element of this term in the  $\mathbf{X}'\mathbf{X}$  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  $\mathbf{X}'\mathbf{X}$  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.

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

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  $\mathbf{X}'\mathbf{X}$ -inverse which is associated with A-optimality.

## D-Optimal Designs

## 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.

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

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

Original Row	Factors		
	A	B	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

## D-Optimal Designs

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

Row	Variable	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

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the **D-Optimal Designs** procedure window by expanding **Design of Experiments**, then clicking on **Experimental Design**, and then clicking on **D-Optimal Designs**. You may then make the appropriate entries as listed below, or open **Example 2** by going to the **File** menu and choosing **Open Example Template**.

#### 1 Specify the design and data storage.

- On the D-Optimal Designs window, select the **Design tab**.
- Set **N Per Block** to **6**.
- Set **Optimize the Design for this Model** to **A|A|B|B**.
- Set **Max Term Order** to **2**.
- Check the **Store Data on Spreadsheet** box.
- Set the **First Factor Column** to **1**.
- Set the **First Expanded Column** to **4**.
- Check **Rename Factor Variables with Factor Labels**.
- Check **Clear Existing Data**.

#### 2 Specify the reports.

- On the D-Optimal Designs window, select the **Reports tab**.
- Set **Decimal Places** to **0**.

#### 3 Run the procedure.

- From the Run menu, select **Start Calculation**. Alternatively, just click the Calculate button.

## 6-Point, 2 Factor D-Optimal Design

A	B
-1	-1
0	-1
1	-1
0	0
-1	1
1	1

Columns A and B give the design. The Determinant Analysis Section showed that the maximum was achieved on 25 of the 30 iterations. Hence, we assume that the algorithm converged to the global maximum.

Next, we 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

A	B
-1	-1
0	-1
0	-1
1	-1
0	0
0	0
-1	1
1	1

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. We call the two new variables A1 and B1. While we are at it, we also create other variables of the expanded design matrix. The resulting database appears as follows.

## 6-Point Design in Expanded Form

A	B	A1	B1	A1_B1	A1_A1	B1_B1
-1	-1	10	1	10	100	1
0	-1	15	1	15	225	1
0	-1	15	1	15	225	1
1	-1	20	1	20	400	1
0	0	15	2	30	225	4
0	0	15	2	30	225	4
-1	1	10	3	30	100	9
1	1	20	3	60	400	9

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

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the **D-Optimal Designs** procedure window by expanding **Design of Experiments**, then clicking on **Experimental Design**, and then clicking on **D-Optimal Designs**. You may then make the appropriate entries as listed below, or open **Example 3** by going to the **File** menu and choosing **Open Example Template**.

#### 1 Specify the design and data storage.

- On the D-Optimal Designs window, select the **Design tab**.
- Set **N Per Block** to **4 4 4**.
- Set **Optimize the Design for this Model** to **A|B|C A\*A B\*B C\*C**.
- Set **Max Term Order** to **2**.
- Set **Max Iterations** to **100**.
- Set **Inclusion Points** to **45**. This is approximately  $(3)(3)(3)(3)/2$  which is the number of blocks times the product of the number of levels in each factor, all divided by two.
- Set **Removal Points** to **11**. This is one less than the total number of points desired.
- Check the **Store Data on Spreadsheet** box.
- Set the **First Factor Column** to **1**.
- Check **Rename Factor Variables with Factor Labels**.
- Check **Clear Existing Data**.

#### 2 Specify the reports.

- On the D-Optimal Designs window, select the **Reports tab**.
- Set **Decimal Places** to **0**.

#### 3 Run the procedure.

- From the Run menu, select **Start Calculation**. Alternatively, just click the Calculate button.

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

A	B	C	Block
-1	-1	-1	2
1	-1	-1	1
-1	0	-1	3
-1	1	-1	1
1	1	-1	2
0	-1	0	3
-1	0	0	2
-1	-1	1	1
1	-1	1	3
0	0	1	2
-1	1	1	3
1	1	1	1

## D-Optimal Designs

Columns A, B, C, and Blocks give the design. The Determinant Analysis Section showed that the maximum was achieved on 12 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

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the **D-Optimal Designs** procedure window by expanding **Design of Experiments**, then clicking on **Experimental Design**, and then clicking on **D-Optimal Designs**. You may then make the appropriate entries as listed below, or open **Example 4** by going to the **File** menu and choosing **Open Example Template**. You can see that the design values are in the spreadsheet by clicking the spreadsheet button on the Design tab.

#### 1 Specify the design and data storage.

- On the D-Optimal Designs window, select the **Design tab**.
- Set **N Per Block** to **4 4 4**.
- Set **Input Columns (Candidate and Forced)** to **A-C**.
- Set **Forced Points** to **8**.
- Set **Optimize the Design for this Model** to **A|B|C A\*A B\*B C\*C**.
- Set **Max Term Order** to **2**.
- Set **Max Iterations** to **30**.
- Set **Inclusion Points** to **5**.
- Set **Removal Points** to **5**.
- Check the **Store Data on Spreadsheet** box.
- Set the **First Factor Column** to **1**.
- Check **Rename Factor Columns with Factor Labels**.
- Check **Clear Existing Data**.

#### 2 Specify the reports.

- On the D-Optimal Designs window, select the **Reports tab**.
- Set **Decimal Places** to **0**.

#### 3 Run the procedure.

- From the Run menu, select **Start Calculation**. Alternatively, just click the Calculate button.



## Augmented D-Optimal Design with Blocking

A	B	C	Blocks
-1	-1	-1	1
1	1	-1	1
1	-1	1	1
-1	1	1	1
1	-1	-1	2
-1	1	-1	2
-1	-1	1	2
1	1	1	2
-1	0	-1	3
0	1	-1	3
1	-1	0	3
0	0	0	3

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

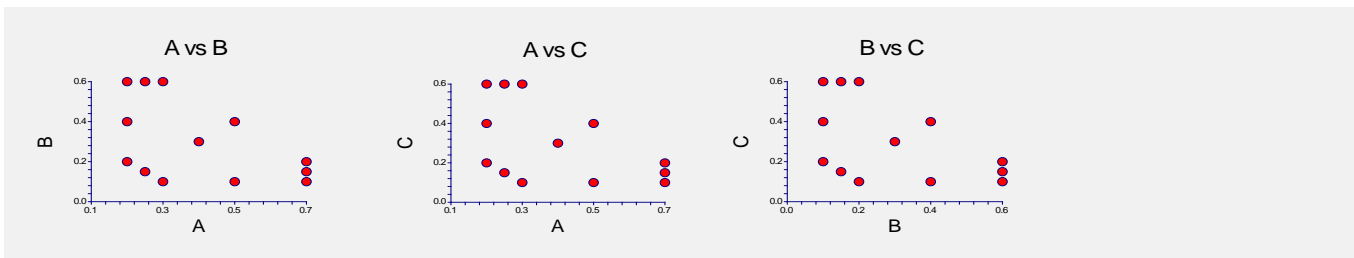
The Determinant Analysis Section showed that the maximum was achieved on 9 of the 30 iterations. Hence, we assume that the algorithm converged to the global maximum.

### 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.

## D-Optimal Designs

## Setup

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the **D-Optimal Designs** procedure window by expanding **Design of Experiments**, then clicking on **Experimental Design**, and then clicking on **D-Optimal Designs**. You may then make the appropriate entries as listed below, or open **Example 5** by going to the **File** menu and choosing **Open Example Template**. You can see that the design values are in the spreadsheet by clicking the spreadsheet button on the Design tab.

### 1 Specify the design and data storage.

- On the D-Optimal Designs window, select the **Design tab**.
- Set **N Per Block** to **10**.
- Set **Input Columns (Candidate and Forced)** to **1-3**.
- Set **Forced Points** to **0**.
- Set **Optimize the Design for this Model** to **A|B|C**.
- Set **Max Term Order** to **2**.
- Set **Max Iterations** to **30**.
- Set **Inclusion Points** to **5**.
- Set **Removal Points** to **5**.
- **Remove the check** from the **Include Intercept** check box.
- Check the **Store Data on Spreadsheet** box.
- Set the **First Factor Column** to **1**.
- Check **Rename Factor Columns with Factor Labels**.
- Check **Clear Existing Data**.

### 2 Specify the reports.

- On the D-Optimal Designs window, select the **Reports tab**.
- Set **Decimal Places** to **4**.

### 3 Run the procedure.

- From the Run menu, select **Start Calculation**. Alternatively, just click the Calculate button.

## Mixture Design

Original Row	A	B	C
1	0.7000	0.1000	0.2000
2	0.2000	0.6000	0.2000
3	0.7000	0.2000	0.1000
4	0.2000	0.2000	0.6000
5	0.3000	0.6000	0.1000
6	0.3000	0.1000	0.6000
8	0.2000	0.4000	0.4000
9	0.5000	0.1000	0.4000
11	0.5000	0.4000	0.1000
13	0.4000	0.3000	0.3000

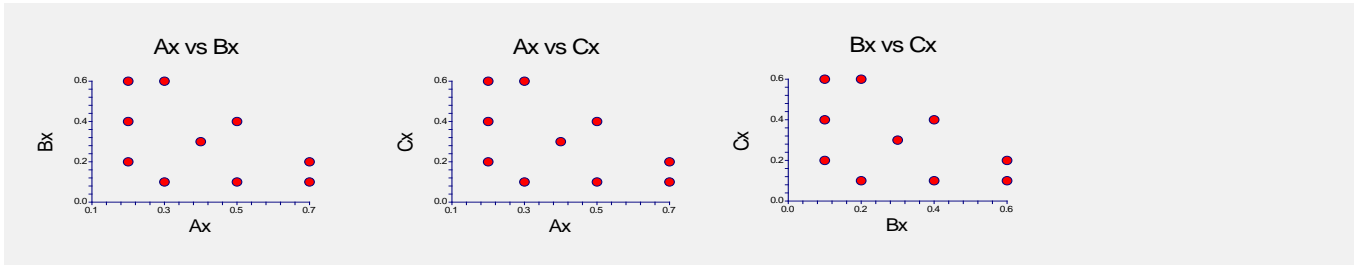
Columns A, B, and C give the design. The original row from the candidate list is shown as the first column of the report.

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

## D-Optimal Designs

## 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

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the **D-Optimal Designs** procedure window by expanding **Design of Experiments**, then clicking on **Experimental Design**, and then clicking on **D-Optimal Designs**. You may then make the appropriate entries as listed below, or open **Example 6** by going to the **File** menu and choosing **Open Example Template**.

#### 1 Specify the data storage.

- On the D-Optimal Designs window, select the **Design tab**.
- Set **N Per Block** to **18**.
- Set **Optimize the Design for this Model** to **A|B|C A\*A B\*B**.
- Set **Max Term Order** to **2**.
- Set **Qualitative Factors and Levels** to **C(5)**.
- Set **Max Iterations** to **30**.
- Set **Inclusion Points** to **20**.
- Set **Removal Points** to **18**.
- **Remove the check** from the **Include Intercept** check box.
- Check the **Store Data on Spreadsheet** box.
- Set the **First Factor Column** to **1**.
- Check **Rename Factor Columns with Factor Labels**.
- Check **Clear Existing Data**.

#### 2 Specify the reports.

- On the D-Optimal Designs window, select the **Reports tab**.
- Set **Decimal Places** to **0**.

#### 3 Run the procedure.

- From the Run menu, select **Run Procedure**. Alternatively, just click the Run button.

## Design with Qualitative Factors

Original Row	Factors		
	A	B	C
1	-1	-1	1
3	1	-1	1
5	0	0	1
9	1	1	1
11	0	-1	2
16	-1	1	2
18	1	1	2
19	-1	-1	3
21	1	-1	3
25	-1	1	3
27	1	1	3
28	-1	-1	4
33	1	0	4
34	-1	1	4
38	0	-1	5
40	-1	0	5
42	1	0	5
44	0	1	5

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.