Chapter 23 Introduction to the OPTEX Procedure

Chapter Table of Contents

OVERVIEW	• •		•	 	• •	· ·					·	719 719 720
GETTING STARTED	• •	· ·	•	 	• •	· ·					. , . ,	721 721 726

Part 6. The CAPABILITY Procedure

Chapter 23 Introduction to the OPTEX Procedure

Overview

The OPTEX procedure searches for optimal experimental designs. You specify a set of candidate design points and a linear model, and the procedure chooses points so that the terms in the model can be estimated as efficiently as possible.

Most experimental situations call for standard designs, such as fractional factorials, orthogonal arrays, or central composites. Standard designs have assured degrees of precision and orthogonality that are important for the exploratory nature of experimentation. In some situations, however, standard designs are not available, such as when

- not all combinations of the factor levels are feasible
- the region of experimentation is irregularly shaped
- resource limitations restrict the number of experiments that can be performed
- there is a nonstandard linear or a nonlinear model

The OPTEX procedure can generate an efficient experimental design for any of these situations.

Note: Instead of using OPTEX directly, a more appropriate tool for you may be the ADX Interface. The ADX Interface, which has been completely revised in Version 7, is designed primarily for engineers and researchers who require a point-and-click solution for the entire experimental process, from building the designs through determining significant effects to optimization and reporting. In addition to offering standard designs as mentioned above, ADX makes it easy to use OPTEX to find optimal designs for non-standard factorial, response surface, and mixture experiments, with and without blocking. Information about the ADX Interface can be found at http://www.sas.com/rnd/app/qc/newadx/newadx.html.

Features

This section summarizes key features of the OPTEX procedure.

The OPTEX procedure offers various criteria for searching a design; these criteria are summarized in Table 23.1 on page 720 and Table 23.2 on page 720. In the formulas for these criteria, X denotes the design matrix, C the set of candidate points, and D the set of design points. The default criterion is D-optimality. You can also use the OPTEX procedure to generate G- and I-efficient designs.

The OPTEX procedure also offers a variety of search algorithms, ranging from a simple sequential search (Dykstra 1971) to the computer-intensive Fedorov algorithm

(Fedorov 1972, Cook and Nachtsheim 1980). You can customize many aspects of the search, such as the initialization method and the number of iterations.

You can use the full general linear modeling facilities of the GLM procedure to specify a model for your design, allowing for general polynomial effects as well as classification or ANOVA effects. Optionally, you can specify

- design points to be optimally augmented
- fixed covariates (for example, blocks) for the design
- prior precisions for Bayesian optimal design

The OPTEX procedure is an interactive procedure. After specifying an initial design, you can submit additional statements without reinvoking the OPTEX procedure. Once you have found a design, you can

- examine the design
- output the design to a data set
- change the model and find another design
- change the characteristics of the search and find another design

Table 23.1. Information-based Optimality Criteria

Criterion	Goal	Formula
D-optimality	Maximize determinant of the	$\max X'X $
	information matrix	
A-optimality	Minimize sum of the variances	min trace $(X'X)^{-1}$
	of estimated coefficients	

Table 23.2. Distance-based Optimality Criteria

Criterion	Goal	Formula
U-optimality	Minimize distance from design to candidates	$\min \sum_{\mathbf{x} \in \mathcal{C}} d(\mathbf{x}, \mathcal{D})$
S-optimality	Maximize distance between design points	$\min \sum_{\mathbf{y} \in \mathcal{D}} d(\mathbf{y}, \mathcal{D} - \mathbf{y})$

Learning about the OPTEX Procedure

To learn the basic syntax of the OPTEX procedure, read the introductory example in the next section, which covers a typical application of optimal designs. Other applications are illustrated in "Optimal Design Scenarios" on page 726. The summary tables in the "Summary of Functions" section on page 732 provides an overview of the syntax. The "Advanced Examples" section on page 744 illustrates construction of complex designs.

Getting Started

The examples in this section illustrate basic features of the OPTEX procedure. In addition, the examples show how a variety of SAS software tools can be used to construct candidate sets. If you are working through these examples on your own computer, note that the randomness in the OPTEX procedure's search algorithm will cause your results to be slightly different from those shown.

See "Advanced Examples" on page 744 for illustrations of complex features.

Constructing a Nonstandard Design

This example shows how you can use the OPTEX procedure to construct a design for a complicated experiment for which no standard design is available.

See OPTEXG1 in the SAS/QC Sample Library

A chemical company is designing a new reaction process. The engineers have isolated the following five factors that might affect the total yield:

Variable	Description	Range
RTEMP	Temperature of the reaction chamber	150-350 degrees
PRESS	Pressure of the reaction chamber	10-30 psi
TIME	Amount of time for the reaction	3-5 minutes
SOLV	Amount of solvent used	20-25 %
SOURCE	Source of raw materials	1, 2, 3, 4, 5

While there are only two solvent levels of interest, the reaction control factors (RTEMP, PRESS, and TIME) may be curvilinearly related to the total yield and thus require three levels in the experiment. The SOURCE factor is categorical with five levels. Additionally, some combinations of the factors are known to be problematic; simultaneously setting all three reaction control factors to their lowest feasible levels will result in worthless sludge, while setting them all to their highest levels can damage the reactor. Standard experimental designs do not apply to this situation.

Creating the Candidate Set

You can use the OPTEX procedure to generate a design for this experiment. The first step in generating an optimal design is to prepare a data set containing the candidate runs (that is, the feasible factor level combinations). In many cases, this step involves the most work. You can use a variety of SAS data manipulation tools to set up the candidate data set. In this example, the candidate runs are all possible combinations of the factor levels except those with all three control factors at their low levels and at their high levels, respectively. The PLAN procedure (refer to the *SAS/STAT User's Guide*) provides an easy way to create a full factorial data set, which can then be subsetted using the DATA step, as shown in the following statements:

```
proc plan ordered;
   factors rtemp=3 press=3 time=3 solv=2 source=5/noprint;
   output out=can
      rtemp nvals=(150 to 350 by 100)
      press nvals=( 10 to 30 by
                                   10)
            nvals=(
                     3 to
                            5
      time
                                     )
      solv
            nvals=( 20 to 25 by
                                    5)
      source nvals=( 1 to
                             5
                                    );
data can; set can;
   if (^((rtemp = 150) & (press = 10) & (time = 3)));
   if (^((rtemp = 350) & (press = 30) & (time = 5)));
proc print data=can;
run;
```

A partial listing of the candidate data set CAN is shown in Figure 23.1.

Obs	rtemp	press	time	solv	source	
1	150	10	4	20	1	
2	150	10	4	20	2	
3	150	10	4	20	3	
4	150	10	4	20	4	
5	150	10	4	20	5	
6	150	10	4	25	1	
7	150	10	4	25	2	
8	150	10	4	25	3	
•	•	•	•	•	•	
•	•	•	•	•	•	
•	•	•	•	•	•	
249	350	30	4	25	4	
250	350	30	4	25	5	



Generating the Design

The next step is to invoke the OPTEX procedure, specifying the candidate data set as the input data set. You must also provide a model for the experiment, using the MODEL statement, which uses the linear modeling syntax of the GLM procedure (refer to the *SAS/STAT User's Guide*). Since SOURCE is a classification factor, you need to specify it in a CLASS statement. To detect possible cross-product effects in the other factors, as well as the quadratic effects of the three reaction control factors, you can use a modified response surface model, as shown in the following statements:

```
proc optex data=can;
    class source;
    model source solv|rtemp|press|time@@2
        rtemp*rtemp press*press time*time;
run;
```

Note that the MODEL statement does not involve a response variable (unlike the MODEL statement in the GLM procedure).

The default number of runs for a design is assumed by the OPTEX procedure to be 10 plus the number of parameters (a total of 10 + 18 = 28 in this case.) Thus, the procedure searches for 28 runs among the candidates in CAN that allow D-optimal estimation of the effects in the model. (See "Optimality Criteria" on page 776 for a precise definition of D-optimality.) Randomness is built into the search algorithm to overcome the problem of local optima, so by default the OPTEX procedure takes 10 random "tries" to find the best design. The output, shown in Figure 23.2, lists efficiency factors for the 10 designs found. These designs are all very close in terms of their D-efficiency.

Design				Average Prediction Standard
Tumber	D-Efficiency	A-Efficiency	G-Efficiency	Error
1	43.6131	22.9676	77.5794	0.8529
2	43.5647	24.8096	77.5054	0.8495
3	43.3119	24.1106	77.8534	0.8494
4	43.0918	23.6364	78.4918	0.8535
5	42.7266	23.5430	75.1467	0.8608
6	42.6481	22.8901	75.6073	0.8669
7	42.5349	22.3007	76.1047	0.8683
8	42.4867	20.7428	77.2003	0.8689
9	42.4342	24.1507	75.5005	0.8562
10	42.1579	20.8750	71.0905	0.8839

Figure 23.2. Efficiency Factors for Chemical Reaction Design

The final step is to save the best design in a data set. You can do this interactively by submitting the OUTPUT statement immediately after the preceding statements. Then use the PRINT procedure to list the design. The design is partially listed in Figure 23.3.

output out=reactor; proc print data=reactor; run;

Obs	solv	rtemp	press	time	source
1	20	150	10	4	5
2	20	150	30	3	5
3	20	350	20	5	5
•	•	•	•	•	•
•	•	•	•	•	•
•	•	•	•	•	•
28	25	350	30	3	1

Figure 23.3. Optimal Design for Chemical Reaction Process Experiment

Customizing the Number of Runs

The OPTEX procedure provides options with which you can customize many aspects of the design optimization process. Suppose the budget for this experiment can only accommodate 25 runs. You can use the N= option in the GENERATE statement to request a design with this number of runs.

```
proc optex data=can;
    class source;
    model source solv|rtemp|press|time@@2
        rtemp*rtemp press*press time*time;
    generate n=25;
run;
```

Including Specific Runs

If there are factor combinations that you want to include in the final design, you can use the OPTEX procedure to *augment* those combinations optimally. For example, suppose you want to force four specific factor combinations to be in the design. If these combinations are saved in a data set, you can force them into the design by specifying the data set with the AUGMENT= option in the GENERATE statement. This technique is demonstrated in the following statements:

```
data preset;
    input solv rtemp press time source;
    datalines;
20 350 10 5 4
20 150 10 4 3
25 150 30 3 3
25 250 10 5 3
;
proc optex data=can;
    class source;
    model source solv|rtemp|press|time@@2
        rtemp*rtemp press*press time*time;
    generate n=25 augment=preset;
    output out=reactor2;
run;
```

The final design is listed in Figure 23.4 on page 725. Note that the points in the AUGMENT= data set appear as observations 7, 11, 15, and 16.

Using an Alternative Search Technique

You can also specify a variety of optimization methods with the GENERATE statement. The default method is relatively fast; while other methods may find better designs, they take longer to run and the improvement is usually only marginal. The method that generally finds the best designs is the modified Fedorov procedure described by Cook and Nachtsheim (1980). The following statements show how to request this method:

```
proc optex data=can;
    class source;
    model source solv|rtemp|press|time@@2
        rtemp*rtemp press*press time*time;
    generate n=25 method=m_fedorov;
run;
```

Obs	solv	rtemp	press	time	source
1	20	150	30	3	5
2	20	350	20	5	5
3	25	150	10	4	5
4	25	250	20	4	5
5	25	350	30	3	5
6	20	150	20	5	4
7	20	250	10	3	4
8	25	150	10	5	4
9	25	150	30	3	4
10	25	350	30	4	4
11	20	150	20	3	3
12	20	250	10	5	3
13	20	350	30	4	3
14	25	150	30	5	3
15	25	350	10	3	3
16	20	150	10	4	2
17	20	250	30	5	2
18	20	350	10	3	2
19	25	250	20	4	2
20	25	350	10	5	2
21	20	150	30	5	1
22	20	350	10	4	1
23	20	350	30	3	1
24	25	150	20	3	1
25	25	250	30	5	1

Figure 23.4. Augmented Design for Chemical Reaction Process Experiment

The efficiencies for the resulting designs are shown in Figure 23.5.

Design Number	D-Efficiency	A-Efficiency	G-Efficiency	Average Prediction Standard Error
1	43.2463	21.5586	76.0488	0.9174
2	43.1226	21.7105	73.4776	0.9171
3	43.1226	21.7105	73.4776	0.9171
4	43.0155	24.7451	73.7483	0.9091
5	42.9533	24.6489	75.3329	0.8977
6	42.8604	23.0747	72.6984	0.9114
7	42.7463	24.8328	78.7114	0.8951
8	42.6978	22.1083	74.1844	0.9219
9	42.6714	21.6200	74.9443	0.9255
10	42.4603	21.6074	74.4040	0.9186



In this case, the modified Fedorov procedure takes three to four times longer than the default method, and D-efficiency only improves by about 0.5%. On the other hand, the longer search method may take only a few seconds on a reasonably fast computer.

Optimal Design Scenarios

The following examples briefly describe some additional common situations that call for optimal designs. These examples show how you can

- use a variety of SAS software tools to generate an appropriate set of candidate runs
- use the OPTEX procedure to search the candidate set for an optimal design

The emphasis here is on the programming techniques; output is omitted.

Constructing a Saturated Second-Order Design

Suppose you want a design for seven two-level factors that is as small as possible but still permits estimation of all main effects and two-factor interactions. Among standard orthogonal arrays, the smallest appropriate 2^k design has 64 runs, far more than the 29 parameters you want to estimate. To generate a D-efficient non-orthogonal design, first use the FACTEX procedure to create the full set of $2^7 = 128$ candidate runs. Then invoke the OPTEX procedure with a full second-order model, asking for a saturated design.

```
proc factex;
  factors x1-x7;
  output out=can1;
proc optex data=can1;
  model x1|x2|x3|x4|x5|x6|x7@@2;
  generate n=saturated;
  output out=design1a;
run;
```

The default search procedure quickly finds a design with a D-efficiency of 82.3%. If search time is not an issue, you can try a more powerful search technique. For example, you can specify 500 tries with the modified Fedorov method.

```
proc optex data=can1;
  model x1|x2|x3|x4|x5|x6|x7@@2;
  generate n=saturated
      method=m_fedorov
      iter=500;
  output out=design1b;
run;
```

This takes more than ten times longer to run, and the resulting design is only slightly more D-efficient.

Augmenting a Resolution 4 Design

In a situation similar to the previous example, suppose you have performed an experiment for seven two-level factors with a 16-run, fractional factorial design of resolution 4. You can estimate all main effects with this design, but some two-factor interactions will be confounded with each other. You now want to add enough runs to estimate all two-factor interactions as well. You can use the FACTEX procedure to create the original design as well as the candidate set.

```
proc factex;
   factors x1-x7;
   output out=can2;
run;
   model resolution=4;
   size design=min;
   output out=aug2;
run;
```

Now specify AUG2 (the data set containing the design to be augmented) with the AUGMENT= option in the GENERATE statement.

```
proc optex data=can2;
model x1|x2|x3|x4|x5|x6|x7@@;
generate n=30 augment=aug2;
output out=design2;
run;
```

Handling Many Variables

When you have many factors, the set of all possible factor level combinations may be too large to work with as a candidate set. Suppose you want a main-effects design for 15 three-level factors. The complete set of $3^{15} > 14,000,000$ candidates is too large to use with the OPTEX procedure; in fact, it will probably be too large to store in your computer. One solution is to find a subset of the full factorial set to use as candidates. For example, an alternative candidate set is the 81-run orthogonal design of resolution 3, which can easily be constructed using the FACTEX procedure.

```
proc factex;
    factors x1-x15 / nlev=3;
    model resolution=3;
    size design=81;
    output out=can3;
proc optex data=can3;
    class x1-x15;
    model x1-x15;
    generate n=saturated;
    output out=design3;
run;
```

Constructing an Incomplete Block Design

An incomplete block design is a design for v qualitative treatments in b blocks of size k, where k < v so that not all treatments can occur in each block. To construct an incomplete block design with the OPTEX procedure, simply create a candidate data set containing a treatment variable with t values and then use the BLOCKS statement. For example, the following statements construct a design for seven treatments in seven blocks of size three:

```
data can4;
    do treatmt = 1 to 7;
        output;
        end;
proc optex data=can4;
    class treatmt;
    model treatmt;
    blocks structure=(7)3;
run;
```

The resulting design is balanced in the sense that each treatment occurs the same number number of times and each pair of treatments occur together in the same number of blocks. Balanced designs, when they exist, are known to be optimal, and the OPTEX procedure usually succeeds at finding them for small- to moderately-sized problems.

Constructing a Mixture-Process Design

Suppose you want to design an experiment with three *mixture factors* X1, X2, and X3 (continuous factors that represent proportions of the components of a mixture) and one *process factor* A (a classification factor with five levels). Furthermore, suppose that X1 can account for no more than 50% of the mixture. You can use the ADXXVERT macro (see page 1896) and the FACTEX procedure (see page 425) to create the candidate set.

```
%adxgen;
%adxmix;
%adxinit;
%adxxvert(xvt,x1 0-.5/x2/x3
proc factex;
   factors a / nlev=5;
   output out=can5 pointrep=xvt;
run;
```

Analyzing mixture designs with linear models can be problematic because of the constraint that the mixture factors sum to one; however, to generate an optimal design, you can simply drop one of the mixture factors. The following statements use the preceding candidate set to find an optimal design for fitting the main effect of A and a second-order model in the mixture factors:

```
proc optex data=can5;
    class a;
    model a x1|x2 x1*x1 x2*x2;
run;
```

See Example 24.10 on page 762 for a more detailed example of a mixture experiment.

The correct bibliographic citation for this manual is as follows: SAS Institute Inc., *SAS/QC[®] User's Guide, Version 8*, Cary, NC: SAS Institute Inc., 1999. 1994 pp.

SAS/QC[®] User's Guide, Version 8

Copyright © 1999 SAS Institute Inc., Cary, NC, USA.

ISBN 1-58025-493-4

All rights reserved. Printed in the United States of America. No part of this publication may be reproduced, stored in a retrieval system, or transmitted, by any form or by any means, electronic, mechanical, photocopying, or otherwise, without the prior written permission of the publisher, SAS Institute Inc.

U.S. Government Restricted Rights Notice. Use, duplication, or disclosure of the software by the government is subject to restrictions as set forth in FAR 52.227–19 Commercial Computer Software-Restricted Rights (June 1987).

SAS Institute Inc., SAS Campus Drive, Cary, North Carolina 27513.

1st printing, October 1999

 SAS^{\circledast} and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute in the USA and other countries. $^{\circledast}$ indicates USA registration.

 $IBM^{\circledast}, ACF/VTAM^{\circledast}, AIX^{\circledast}, APPN^{\circledast}, MVS/ESA^{\circledast}, OS/2^{\circledast}, OS/390^{\circledast}, VM/ESA^{\circledast}, and VTAM^{\circledast}$ are registered trademarks or trademarks of International Business Machines Corporation. $^{\circledast}$ indicates USA registration.

Other brand and product names are registered trademarks or trademarks of their respective companies.

The Institute is a private company devoted to the support and further development of its software and related services.