Paper 196-2010

Custom Designs Using JMP® Design of Experiments and SAS® PROC OPTEX

Mei-Fen Yeh, Anthony Cece, Mark Presser
Unilever, Trumbull, CT

ABSTRACT

A proper experimental design is important to ensure that data collected in a study is relevant to answer research questions. However, a predefined, textbook classical design rarely provides an exact match to real-world research problems. There are usually unique restrictions and constraints involved in the whole experimentation process. Fortunately, with the advancement of computing power and SAS® software development, users can generate custom designs based on optimal algorithms. In this paper, we present a flexible approach using JMP® Design of Experiments (DOE) designer in conjunction with the SAS OPTEX procedure to create optimal experiments that are tailored to our specific needs. We illustrate our approach in the context of formulation optimization. The discussion is focused on how to generate different combinations of formulations in a systematic way, and how to assign treatments to experimental units in a way that minimizes experimental error. The creation and evaluation of such a design is discussed.

MOTIVATING EXAMPLE

Formulation optimization is a critical process for R&D in the Fast Moving Consumer Goods Industry (FMCG). It is desirable to gain a better understanding of interactions between formulation components, and how these formulations impact consumers. Consider the following motivating example. A formulator is interested in measuring the effects of four solvents *in vivo*. From historical data and literature reviews, these solvents are known to dissolve a particular family of biological actives. Each of these four solvents has different lipophilic and penetrating properties. In this particular type of *in vivo* study, consumer subjects are treated on the arms with a variety of solvent mixtures. The response variable to be measured is irritation experience. The study objective is to build predictions of the response for any mixture or combination of the ingredients within the experimental range. The results will be used to select optimal formulations providing maximal delivery of actives without inducing undesirable irritation experiences in consumers.

There are several challenges associated with our particular research problem. The formulation components are subject to the constraint that the total proportion of ingredient levels must sum to one. It is of interest to evaluate the relative proportion of the formulation ingredients rather than the actual quantity. In addition, there are often practical limitations resulting in a maximum and/or minimum value for each component.

Another challenge involves assigning solvent mixtures to test subjects. The restriction is that only a limited number of mixtures can be tested on each subject. Besides, there are other factors which are known or suspected to impact the outcome variable. While it is desirable to test solvent mixtures on all possible combinations of experimental factors, it is usually impractical and costly. We need to obtain a balanced design which reduces the experimental error and leads to more precise estimates of the formulation effects.

In the next two sections, we present a flexible approach using JMP Design of Experiments (DOE) designer in conjunction with the SAS OPTEX procedure to create experiments that fit specific research purposes. We illustrate our approach in the context of formulation optimization. The discussion is focused on (1) how to generate different combinations of formulations in a systematic way, and (2) how to assign different combinations of solvent mixtures to consumer subjects in a way that minimizes experimental error.

JMP DESIGN OF EXPERIMENTS (DOE) FOR MIXTURE DESIGNS

In the first step of the experimentation, we want to obtain a treatment structure which creates combinations of levels of formulations in a systematic way. From historical data and literature reviews, four solvents are known to dissolve a particular family of biological actives. Each of these solvents has different lipophilic and penetrating properties: 1)

solvent A: a versatile solubilizer, penetration enhancer for actives which dissolve well in it; 2) solvent B: a co-solvent which works together with solvent A to enhance penetration; 3) solvent C: a non-irritating, non-penetrating polar polymer making up the balance of the formulation; 4) solvent D: a slightly penetrating, less polar polymer alternative to solvent C. The individual solvents are restricted by the following ranges: $0.1 \le A \le 0.5$, $0.1 \le B \le 0.35$, $0 \le C \le 0.8$ and $0 \le D \le 0.8$. We are interested in building a predictive model with all four solvents as main effects and their associated second order interaction terms.

A mixture design is useful to create experiments where the response depends on mixtures of ingredients. Mixture experiments are often encountered in product formulation problems which involve combining two or more ingredients together. A detailed discussion of experiments with mixtures can be found in Cornell (2002), but a simple description follows. The key characteristic of mixture designs is that the total proportion of ingredient levels must sum to one. This constraint complicates the design since ingredients can not vary independently as in factorial designs. When the proportion of one component changes, the proportion of one or other components must change. The problem becomes more complex, when additional restrictions apply, such as maximum or minimum values for each component.

There are several classical mixture design approaches, such as simplex centroid designs, simplex lattice designs, and extreme vertices designs. Simplex centroid designs and simplex lattice designs are not appropriate in our situation because both methods require all the factors range from 0 to 100%. The extreme vertices design accounts for range limits, but is not necessary optimal when additional model structure is assumed. With the capability of JMP Design of Experiments, we do not need to limit ourselves to pre-existing standard designs. We can let the computer find appropriate designs given specific inputs and model structures.

In JMP, one can create mixture experiments by using the mixture designer or the custom designer. We use our motivating example to illustrate how to create mixture designs by employing the custom designer (**DOE > Custom Design)**. Figure 1 below shows the inputs. The factor panel captures solvent factors and their associated minimum/maximum values. The model panel describes a model with main effects and interaction terms. This model allows us to account for synergic or antagonistic effects of pairs of components. It is important to note that in mixture designs, quadratic terms are confounded with interaction terms. As a result, an error will be given if the user is trying to include second order power terms.

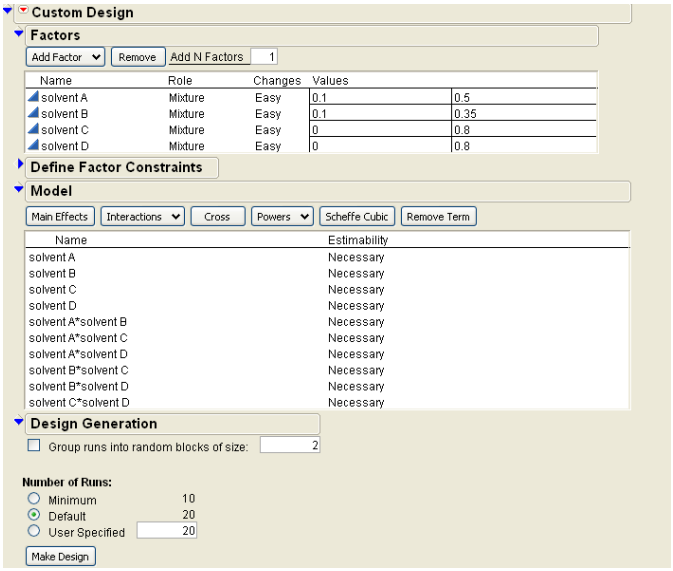


Figure 1 Custom Design Input

Alternatively, one can create optimal designs by invoking **DOE > Mixture Design** and then choose **Optimal** tab after all the factors are entered. By default, JMP generates a 20-run design. The resulting design depends on a random number seed. Figure 2 describes the design generated.

Custom Design

Design

Run	solvent A	solvent B	solvent C	solvent D	Y
1	0.26	0.1	0.32	0.32	.
2	0.3	0.225	0	0.475	.
3	0.1	0.1	0	0.8	.
4	0.5	0.1	0	0.4	.
5	0.1	0.225	0.3375	0.3375	.
6	0.3	0.1	0.3	0.3	.
7	0.1	0.1	0.8	0	.
8	0.1	0.35	0.55	0	.
9	0.5	0.25	0	0.25	.
10	0.5	0.35	0.15	0	.
11	0.1	0.225	0.3375	0.3375	.
12	0.3	0.35	0	0.35	.
13	0.325	0.225	0.45	0	.
14	0.1	0.35	0	0.55	.
15	0.1	0.1	0.4	0.4	.
16	0.3	0.35	0.35	0	.
17	0.3	0.2125	0	0.4875	.
18	0.325	0.2	0.475	0	.
19	0.5	0.1	0.4	0	.
20	0.26	0.225	0.309	0.206	.

Design Evaluation

Output Options

Run Order: Randomize

Make JMP Table from design plus

Number of Center Points: 0

Number of Replicates: 0

Make Table

Back

Figure 2 Custom Design for a Mixture Experiment

TERNARY PLOTS

Once the design is created, one can visualize the geometric representation by a ternary plot. To create ternary plots in JMP, select **Graph > Ternary Plot.**

One should exercise caution when trying to visualize more than three components. In our example, four components are needed to make the formulation. Suppose we choose solvents A, B, and C as the **X, Plotting** variables. The resulting plot is shown in Figure 3.1. The highlighted data point corresponds to the 15th run in the design table. As indicated in Figure 3.2, the sum of solvents A, B, and C is equal to 0.6. JMP then rescales each of these values so that these three components add up to one. The rescaled values (0.1/0.6, 0.1/0.6, and 0.4/0.6), are plotted in the ternary plot.

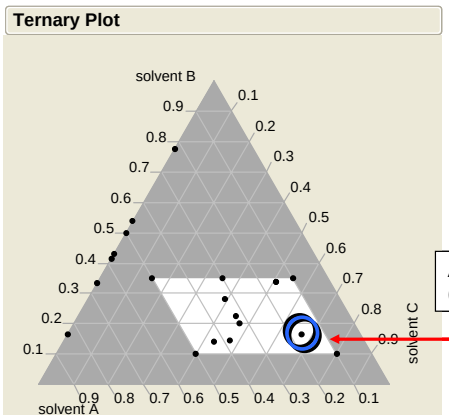
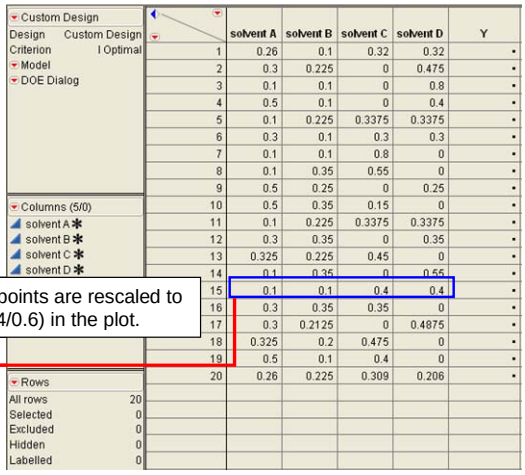


Figure 3.1



Actual design data points are rescaled to (0.1/0.6, 0.1/0.6, 0.4/0.6) in the plot.

Figure 3.2

Figure 3.1 Ternary Plot of Three Components with Rescaled Values
Figure 3.2 Design Table

To see ternary plots that correctly represent the values in this 4 dimensional case, one would need to select all 4 solvents as **X, Plotting** variables. JMP will then produce a series of ternary plots that show two variables and the sum of the other two variables. These plots will reflect the actual values of the design points.

Another way to create ternary plots showing actual design points is to graph three components by various levels of the components not shown on the plot. For example, we can choose solvents B, C, and D as the **X, Plotting** variables, and solvent A as the **by** variable. Figure 4 shows one of the resulting plots.

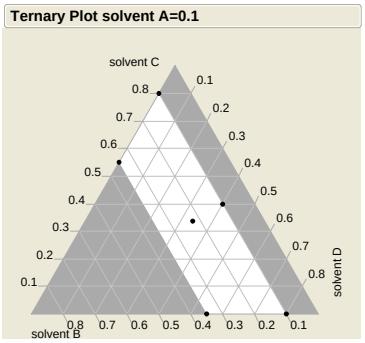


Figure 4 Ternary Plot of Solvents B, C, and D, Given Solvent A Proportion Fixed at 0.1

OPTIMALITY CHOICES

The custom designer generates designs based on two different types of optimality criteria: *D*-Optimal and *I*-Optimal criteria. The *I*-Optimal design is appropriate for our motivating example where the primary goal is to build a predictive model for formulation optimization. This type of design seeks to minimize the average variance of predictor over the design region. *D*-Optimal designs are helpful for generating screening experiments where precise estimates of the coefficients are desired. For comparison, we generate designs by two different optimality criteria, based on the exact same factors and model specification. The design specification is described in Figure 1. In the following paragraphs, we illustrate how to evaluate and compare these two resulting designs using visualization tools.

Ternary plots displayed in Figure 5 allow one to visualize how design points are allocated in the design space. It is easy to see that the *I*-Optimal design tends to assign more design points at the inner design space, while the *D*-Optimal design pushes more design points at the corners.

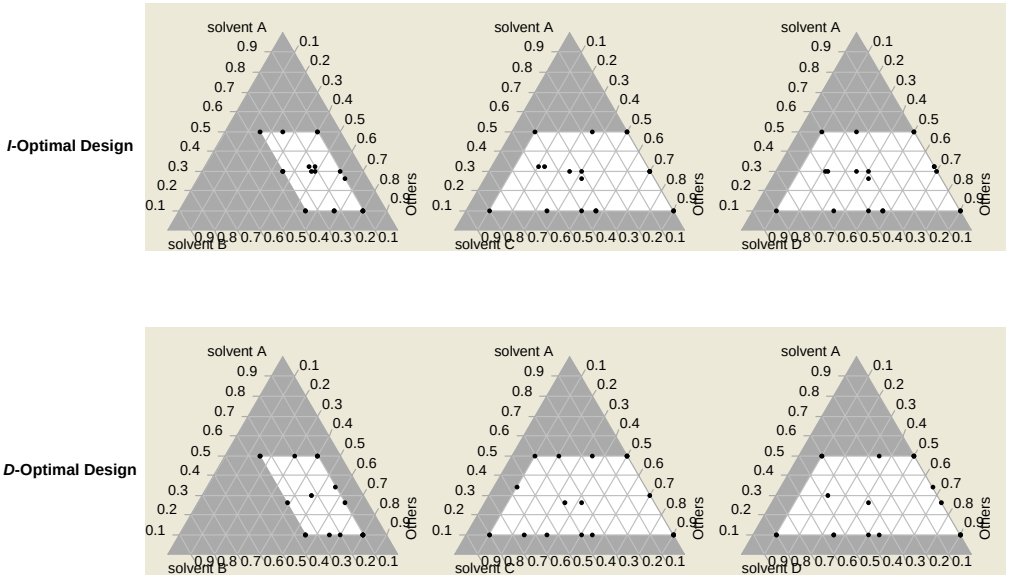


Figure 5 Partial Ternary Plots for *I*-Optimal and *D*-Optimal Designs

A prediction variance profile plot is useful to visualize the variance as a function of each design point. It is shown in Figure 6 that the *I*-Optimal design has the lowest prediction variance at the center of the design space. The price of this characteristic is the higher uncertainty of prediction at the extremes.

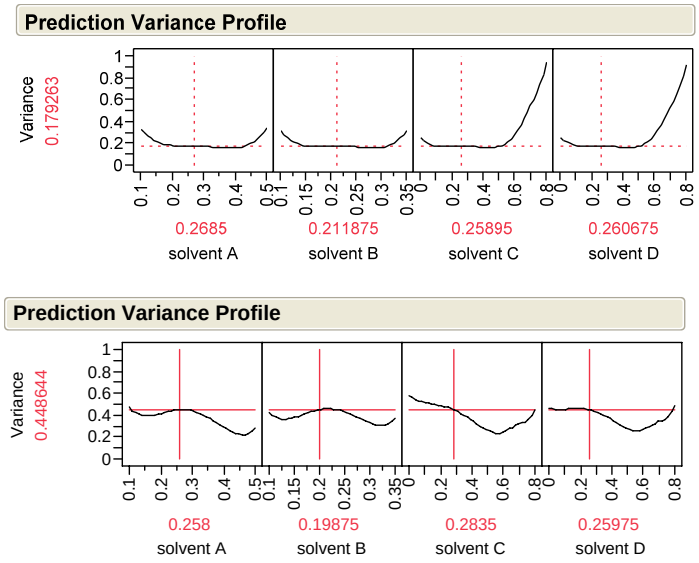


Figure 6 Prediction Variance Profile Plots for *I*-Optimal and *D*-Optimal Designs

The average prediction variance can be found under the Design Diagnostics section. The average variance of the *I*-Optimal design is 0.26, compared to 0.37 from the *D*-Optimal design. As a result, the *I*-Optimal design is appropriate for predicting the response inside the design region. JMP also provides an interactive way to visualize the prediction variance over the design space using a fraction of design space plot. The plot on the left in Figure 7 indicates that in the *I*-Optimal design, 50% of the prediction variance values are smaller than 0.26.

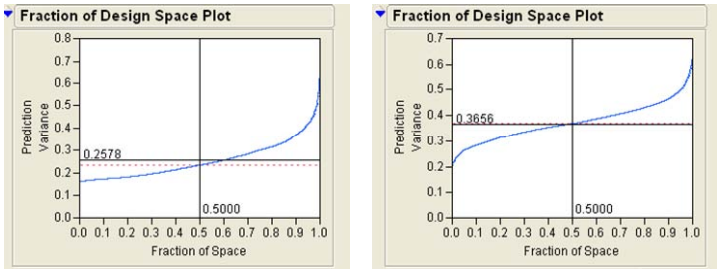


Figure 7 Fraction of Design Space Plots for *I*-Optimal (Left) and *D*-Optimal (Right) Designs

AUGMENTED DESIGNS

The augment designer is a useful tool to modify designs. One can add replicates or center points by choosing these options from available tabs in the augment designer panel. It also provides a flexible way which allows users to make any changes to the design table. The users can then examine competing designs by diagnostic statistics to see the impact of changing those particular design points.

Suppose a researcher is interested in combining the resulting *I*-Optimal design with additional design points chosen from an extreme vertices design. To create an augmented design and evaluate the resulting design,

- Copy additional design points to the existing design table. The modified design table is shown in Figure 8. The labeled rows represent additional runs.
- Invoke **DOE > Augment Design.**
- Select **augment** button under the **augmentation options**.

- Enter 24 under the **Design Generation** box, and click **Make Table**.
- Click **Design Diagnostics** to evaluate the augmented design.

	solvent A	solvent B	solvent C	solvent D	Y
1	0.26	0.1	0.32	0.32	.
2	0.3	0.225	0	0.475	.
3	0.1	0.1	0	0.8	.
4	0.5	0.1	0	0.4	.
5	0.1	0.225	0.3375	0.3375	.
6	0.3	0.1	0.3	0.3	.
7	0.1	0.1	0.8	0	.
8	0.1	0.35	0.55	0	.
9	0.5	0.25	0	0.25	.
10	0.5	0.35	0.15	0	.
11	0.1	0.225	0.3375	0.3375	.
12	0.3	0.35	0	0.35	.
13	0.325	0.225	0.45	0	.
14	0.1	0.35	0	0.55	.
15	0.1	0.1	0.4	0.4	.
16	0.3	0.35	0.35	0	.
17	0.3	0.2125	0	0.4875	.
18	0.325	0.2	0.475	0	.
19	0.5	0.1	0.4	0	.
20	0.26	0.225	0.309	0.206	.
21	0.3	0.35	0.35	0	.
22	0.3	0.225	0.475	0	.
23	0.3	0.1	0.3	0.3	.
24	0.3	0.35	0.175	0.175	.

Figure 8 Design Augmented with Extra Runs Chosen from an Extreme Vertices Design

PROC OPTEX FOR BALANCING TREATMENTS TO SUBJECTS AND APPICATION SITES

In the next step, a treatment allocation scheme is needed to assign different combinations of solvent mixtures to consumer subjects.

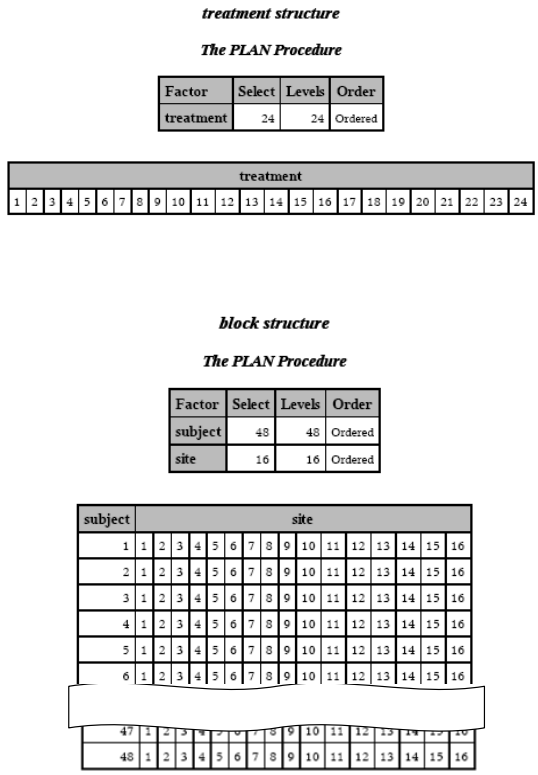
Blocking is often applied in experimental designs as a method of reducing experimental error. It is achieved by assigning treatments to homogeneous experimental units (blocks). In a complete block design, every treatment is allocated in every block. However, complete block designs are not always possible especially when there are a lot of treatments. Suppose we want to assign twenty four different solvent mixtures (treatments) to a total of forty eight subjects. The restriction is that only a limited number of mixtures can be tested on each subject. There are only sixteen application sites available for each subject (block). This is an incomplete block design where the number of treatments (*v*) is larger than the number of runs (*k*) in each block (*b*).

A balanced incomplete block design (BIBD) was introduced by Yates (1936), which guarantees that any given two treatments assigned in the same block (subject) are the same for all pairs. In our situation, a BIBD allows us to account for subject variability. In addition, there are other factors which are known or suspected to impact the outcome variable. For example, the application site effect is considered to contribute to the variability observed in the response data. In this context, the subject and site effects represent two sources of variation. One approach is to construct a Youden Square design (Cochran and Cox, 1957), a special class of Row-Column designs which controls two sources of variation simultaneously. However, mathematically, not all combinations of *v,b,k* result in a BIBD or a Youden design. Fortunately, using the SAS OPTEX procedure, we can generate an optimal, non-standard experimental design with any combination of treatments, application sites, and subjects.

The following program illustrates how to generate a balanced design. The idea is to take twenty four treatments and assign them to forty eight subjects, each of whom has sixteen available application sites. The PLAN procedure creates the treatment and block structure. The resulting SAS output is shown in Figure 9.

```
/* Set up treatment structure: a total of 24 treatments
/-----------------------------------------------------*/
title "treatment structure";
proc plan;
  factors treatment=24 ORDERED ;
  output out=treat_24;
run;

/* Set up block structure: a total of 48 subjects, each subject has 16 application
sites
/-----------------------------------------------------*/
title "block structure";
proc plan;
  factors subject=48 ORDERED site=16 ORDERED ;
  output out=block_48;
run;
```



treatment structure

The PLAN Procedure

Factor	Select	Levels	Order
treatment	24	24	Ordered

treatment																							
1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24

block structure

The PLAN Procedure

Factor	Select	Levels	Order
subject	48	48	Ordered
site	16	16	Ordered

subject	site														
1	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
2	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
3	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
4	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
5	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
6	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
47	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
48	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16

Figure 9 PROC PLAN Outputs for Treatment Structure (Top) and Block Structure (Bottom)

We then take the resulting treatment and block structure as inputs for the OPTEX procedure to find an optimal design. In the following code, the first MODEL statement defines the treatment model. The second MODEL statement that follows the BLOCKS statement, describes the block effect. The partial resulting design is displayed in Figure 10.

```
/ Find the optimal design.
/-----------------------------------------------------*/

proc optex data=treat_24;
   class Treatment;
   model Treatment;
   blocks design=block_48;
   class Subject site;
   model Subject site;
   output out=Design;
run;
```

optimal design

Obs	subject	site	treatment
1	1	1	1
2	1	2	4
3	1	3	21
4	1	4	7
5	1	5	22
6	1	6	5
7	1	7	8
8	1	8	11
9	1	9	2
10	1	10	10
11	1	11	18
12	1	12	15
13	1	13	19
14	1	14	24
15	1	15	12
16	1	16	9
17	2	1	4
18	2	2	11
19	2	3	15
20	2	4	16
21	2	5	7

Figure 10 Partial Listing of the Final Design Table

One can, by using *D*-efficiency, or other efficiency statistics evaluate the design relative to hypothetical balanced designs which might not exist. Here we illustrate a simple way to visualize the resulting design. The tables displayed in Figure 11 are coded using additional steps not shown here. It is displayed in Figure 11.1 that treatment 1 and 2 appear together in 21 of the subjects, treatment 2 and 3 appear together in 21 of the subjects, etc. In this design, there is a total of 276 pairwise combinations (twenty four choose two). In most cases, any given pair of treatments appears together in 20-22 subjects. It also shows that each treatment is replicated 32 times. Figure 11.2 indicates that the design is perfectly balanced with respect to treatment assignments by site (site 1 through site 16). In other words, treatment is assigned in the first application site 2 times, in the second application site 2 times, etc. In summary, each pair of treatments gets a similar chance of being assigned to the same subject, and each application site gets the same chance of being exposed to each treatment. This ensures that treatments are allocated to similar conditions to minimize the effects other than treatment interventions.

treatment combination	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24
1	32	21	21	21	20	21	21	20	21	20	21	21	21	21	22	20	22	20	21	22	21	21	21	20
2	21	32	21	21	20	21	21	22	20	21	20	21	20	20	20	21	21	21	21	22	20	21	22	22
3	21	21	32	21	21	20	21	22	20	21	21	21	21	21	21	22	21	20	20	21	20	20	21	22
4	21	21	21	32	21	21	22	20	20	20	20	20	21	21	21	22	20	21	20	21	22	22	21	21
5	20	20	21	21	32	21	21	21	22	21	20	20	21	21	20	22	21	21	21	21	21	22	21	20
6	21	21	20	21	21	32	21	22	20	21	21	21	21	20	20	22	20	21	20	21	21	22	21	21
7	21	21	21	22	21	21	32	20	21	20	21	21	21	22	21	20	22	21	20	21	22	19	20	21
8	20	22	22	20	21	22	20	32	22	20	21	21	21	21	21	21	20	21	21	20	21	21	21	20
9	21	20	20	20	22	20	21	22	32	22	21	20	20	21	21	21	20	22	22	21	21	20	21	21
10	20	21	21	20	21	21	20	20	22	32	21	21	21	22	22	21	21	21	21	21	21	22	20	19
11	21	20	21	20	20	21	21	21	21	21	32	21	21	21	21	21	21	22	21	19	22	20	21	22
12	21	21	21	20	20	21	21	21	20	21	21	32	21	21	21	21	21	20	21	20	21	23	22	20
13	21	20	21	21	21	21	21	21	20	21	21	21	32	21	21	20	21	21	22	21	21	20	21	21
14	21	20	21	21	21	20	22	21	21	22	21	21	21	32	21	20	21	21	20	20	21	21	20	22
15	22	20	21	21	20	20	21	21	21	22	21	21	21	21	32	21	21	20	21	21	20	22	20	21
16	20	21	22	22	22	22	20	21	21	21	21	21	20	20	21	32	20	20	21	21	21	20	21	21
17	22	21	21	20	21	20	22	20	20	21	21	21	21	21	21	20	32	21	21	21	21	21	20	22
18	20	21	20	21	21	21	21	21	22	21	22	20	21	21	20	20	21	32	21	22	20	21	21	21
19	21	21	20	20	21	20	20	21	22	21	21	21	22	20	21	21	21	21	32	22	21	20	21	21
20	22	22	21	21	21	21	20	21	21	19	20	21	20	21	21	21	22	22	32	20	20	22	20	
21	21	20	20	22	21	21	22	21	21	21	22	21	21	21	20	21	21	20	21	20	32	20	21	21
22	21	21	20	22	22	22	19	21	20	22	20	23	20	21	22	20	21	21	20	20	20	32	21	21
23	21	22	21	21	21	21	20	21	21	20	21	22	21	20	20	21	20	21	21	22	21	21	32	20
24	20	22	22	21	20	21	21	20	21	19	22	20	21	22	21	21	22	21	21	20	21	21	20	32

Figure 11.1 Frequency Table of Pairwise Treatment Occurrences within Subject

treatment \ site	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
1	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2
2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2
3	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2
4	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2
5	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2
6	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2
7	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2
8	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2
9	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2
10	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2
11	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2
12	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2
13	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2
14	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2
15	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2
16	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2
17	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2
18	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2
19	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2
20	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2
21	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2
22	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2
23	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2
24	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2

Figure 11.2 Frequency Table of Treatments and Application Sites

CONCLUSION

In the real world, each research problem has its unique objective, process, and constraints. A classical textbook experiment design rarely provides a perfect match that suits the specific need. In this paper, we illustrate a flexible approach using JMP DOE designer in conjunction with the SAS OPTEX procedure to create custom experiments. While our discussion is focused on the application of formulation optimization, the same approach could be successfully applied to other areas of applications.

REFERENCES

Cochran, W.G. and Cox, G.M. (1957), *Experimental Designs, Second Edition,* New York: John Wiley & Sons, Inc.

Cornell, J.A. (2002), *Experiments with Mixtures: Designs, Models, and the Analysis of Mixture Data*, New York: John Wiley and Sons, Inc.

Yates, F. (1936), "Incomplete Randomized Blocks," *Annals of Eugenics*, 7, 121 - 140.

SAS Institute Inc. 2009. *JMP® 8 Design of Experiments Guide, Second Edition.* Cary, NC: SAS Institute Inc.

SAS Institute Inc. 2008. *SAS/QC® 9.2 User's Guide.* Cary, NC: SAS Institute Inc.

SAS Institute Inc. 2009. *JMP® 8 Statistics and Graphics Guide, Second Edition.* Cary, NC: SAS Institute Inc.

ACKNOWLEDGMENTS

We greatly appreciate Kathleen Kiernan, SAS Technical Support Team, who assisted us with developing the SAS OPTEX program.

CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the author at:

Mei-Fen Yeh
Unilever
40 Merritt Blvd
Trumbull, CT 06611
Phone: 203-381-2592
mei-fen.yeh@unilever.com