Design of Experiments Example: A Supersaturated Design

It is common for brainstorming sessions to identify dozens of potentially active factors. Rather than reduce the list without the benefit of data, you can use a supersaturated design.

In a *saturated* design, the number of runs equals the number of model terms. In a *supersaturated* design, the number of model terms exceeds the number of runs (Lin, 1993). A supersaturated design can examine dozens of factors using fewer than half as many runs as factors. This makes it an attractive choice for factor screening when there are many factors and experimental runs are expensive.

Limitations of Supersaturated Designs

There are drawbacks to supersaturated designs:

- If the number of *active* factors is more than half the number of runs in the experiment, then it is likely that these factors will be impossible to identify. A general rule is that the number of runs should be at least four times larger than the number of active factors. In other words, if you expect that there might be as many as five active factors, you should plan on at least 20 runs.
- Analysis of supersaturated designs cannot yet be reduced to an automatic procedure. However, using forward stepwise regression is reasonable. In addition, the Screening platform (**Analyze** > **Modeling** > **Screening**) offers a streamlined analysis.

Generate a Supersaturated Design

In this example, you want to construct a supersaturated design to study 12 factors in 8 runs. To create a supersaturated design, you set the Estimability of all model terms (except the intercept) to If Possible.

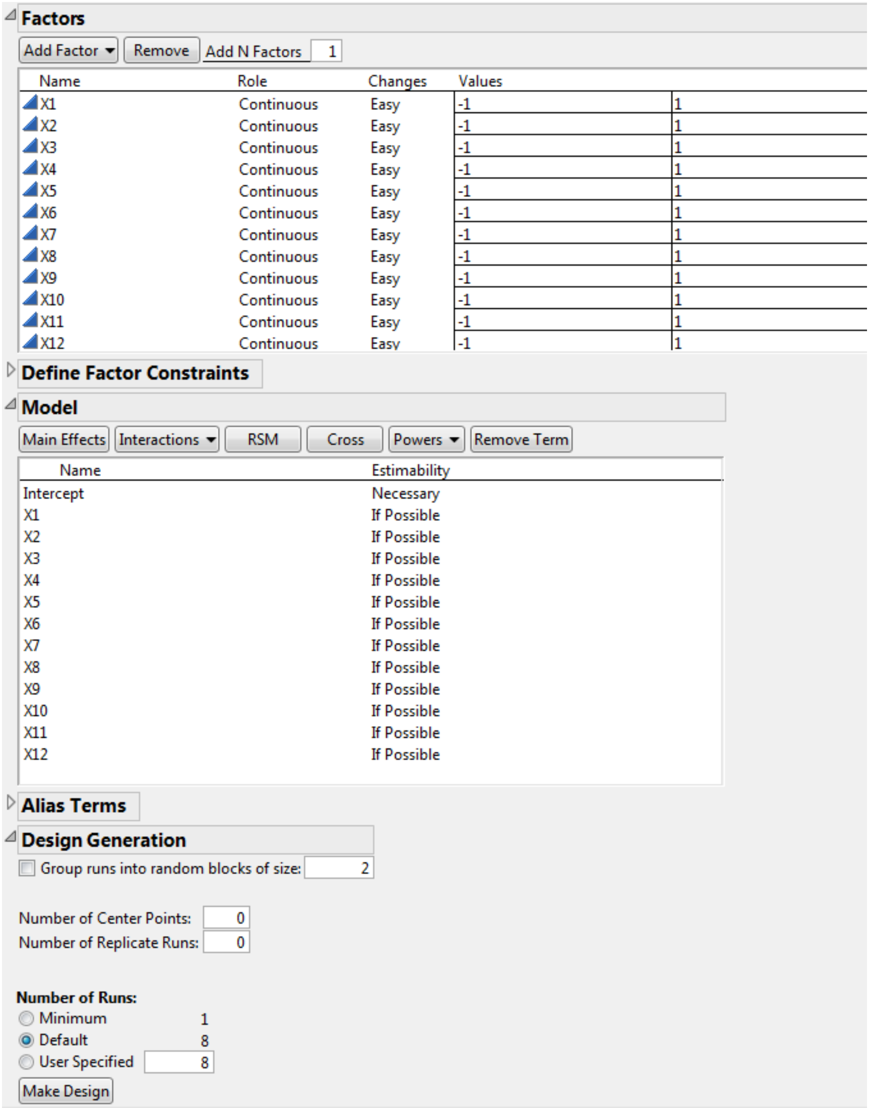
Note: This example is for illustration only. You should have at least 14 runs in any supersaturated design. If there are as many as four active factors, it is very difficult to interpret the results of an 8-run design.

---

1. Select **DOE > Custom Design**.
2. Type 12 next to **Add N Factors**.
3. Click **Add Factor > Continuous**.
4. Click **Continue**.
5. In the Model outline, select all terms except the Intercept.
6. Click Necessary next to any effect and change it to If Possible.

   Setting the effects to If Possible ensures that JMP uses the Bayesian D-optimality criterion to obtain the design.

Figure 1 Factors, Model, and Number of Runs



7. In the Alias Terms outline, select all effects and click **Remove Term**.

   This ensures that only the main effects appear in the Color Map on Correlations. This plot is constructed once the design is created.

8. Select **Simulate Responses** from the red triangle menu.

   This option generates random responses that appear in your design table. You will use these responses to see how to analyze experimental data.

   Keep the Number of Runs set to the Default of 8.

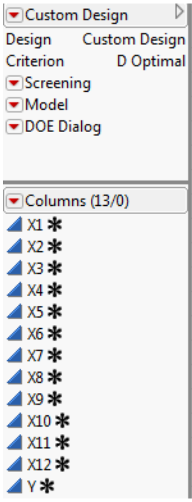
Note: Setting the Random Seed in step 9 and Number of Starts in step 10 reproduces the exact results shown in this example. In constructing a design on your own, these steps are not necessary.

9. (Optional) From the Custom Design red triangle menu, select **Set Random Seed**, type 1008705125, and click **OK**.
10. (Optional) From the Custom Design red triangle menu, select **Number of Starts**, type 100, and click **OK**.
11. Click **Make Design**.
12. Click **Make Table**.

Do not close your Custom Design window. You return to it later in this example.

The design table (Figure 2) and the Simulate Responses window (Figure 3) appear.

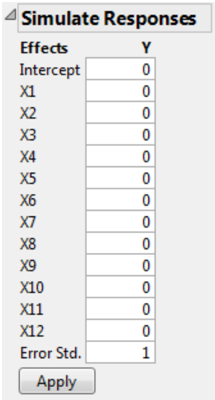
Figure 2 Design Table with Simulated Responses



	X1	X2	X3	X4	X5	X6	X7	X8	X9	X10	X11	X12	Y
1	-1	-1	1	1	-1	1	-1	-1	-1	1	-1	-1	0.4464310057
2	1	1	1	-1	-1	-1	-1	-1	1	-1	1	-1	0.6191292817
3	1	1	-1	1	1	-1	-1	1	-1	1	-1	-1	1.3067571042
4	1	-1	1	-1	1	1	-1	1	1	-1	-1	1	-0.959675933
5	-1	1	-1	1	1	1	1	-1	1	-1	-1	-1	-1.090307891
6	-1	1	1	1	-1	-1	1	1	-1	-1	1	1	0.9802721961
7	-1	-1	-1	-1	-1	-1	1	1	1	1	-1	-1	-1.425139951
8	1	-1	-1	-1	1	1	1	-1	-1	1	1	1	0.1255596767

The response column, Y, contains simulated values. These are randomly generated using the model defined by the parameter values in the Simulate Responses window.

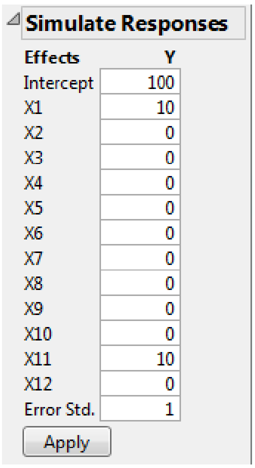
Figure 3 Simulate Responses Window



The Simulate Responses window shows coefficients of 0 for all terms, with an Error Std of 1. The values in the Y column currently reflect only random variation. Notice that the model coefficients are set to 0 because not all coefficients are estimable.

13. Change the values of the coefficients in the Simulate Responses window as shown in Figure 4.

Figure 4 Parameter Values for Simulated Responses

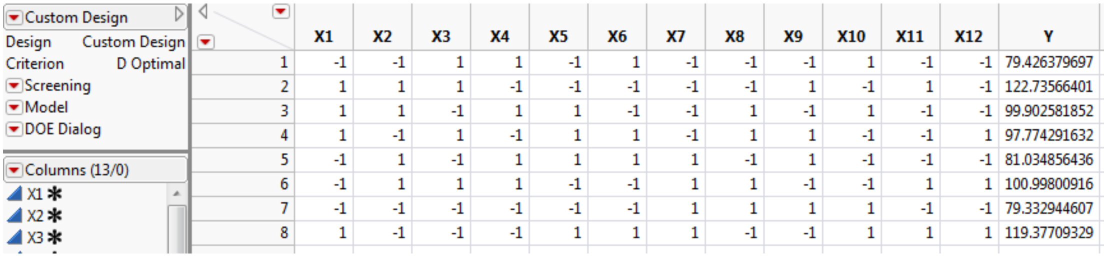


14. Click **Apply**.

The response values in the Y column change. See Figure 5.

Note: If you did not set the random seed and the number of starts, or if you click Apply more than once, your response values will not match those in Figure 5.

Figure 5 Response Column with X1 and X11 Active


Custom Design
Design Custom Design
Criterion D Optimal
Screening
Model
DOE Dialog
Columns (13/0)
X1
X2
X3

	X1	X2	X3	X4	X5	X6	X7	X8	X9	X10	X11	X12	Y
1	-1	-1	1	1	-1	1	-1	-1	-1	1	-1	-1	79.426379697
2	1	1	1	-1	-1	-1	-1	-1	1	-1	1	-1	122.73566401
3	1	1	-1	1	1	-1	-1	1	-1	1	-1	-1	99.902581852
4	1	-1	1	-1	1	1	-1	1	1	-1	-1	1	97.774291632
5	-1	1	-1	1	1	1	1	-1	1	-1	-1	-1	81.034856436
6	-1	1	1	1	-1	-1	1	1	-1	-1	1	1	100.99800916
7	-1	-1	-1	-1	-1	-1	1	1	1	1	-1	-1	79.332944607
8	1	-1	-1	-1	1	1	1	-1	-1	1	1	1	119.37709329

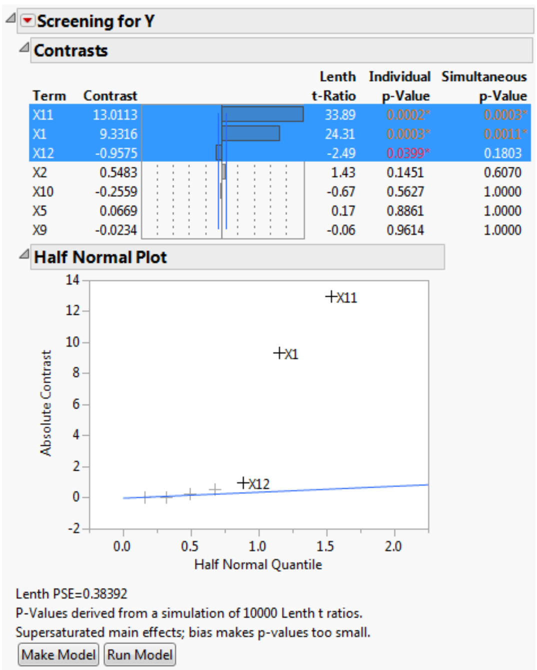
In your simulation, you specified X1 and X11 as active factors with large effects relative to the error variation. For this reason, your analysis of the data should identify these two factors as active.

Analyze a Supersaturated Design Using the Screening Platform

The Screening platform provides a way to identify active factors. The design table in Figure 5 contains three scripts. The Screening script analyzes your data using the Screening platform (located under the **Analyze > Modeling > Screening** menu).

1. In the Tables panel of the design table, select **Run Script** from the red triangle next to **Screening**.

Figure 6 Screening Report for Supersaturated Design



The factors X1 and X11 have large contrast and Lenth t-Ratio values. Also, their Simultaneous p-Values are small. In the Half Normal Plot, both X1 and X11 fall far from the line. The Contrasts and the Half Normal Plot reports indicate that X1 and X11 are active. Although X12 has an Individual p-Value less than 0.05, its effect is much smaller than that of X1 and X11.

Because the design is supersaturated, *p*-values might be smaller than they would be in a model where all effects are estimable. This is because effect estimates are biased by other potentially active main effects. In Figure 6, a note directly above the Make Model button warns you of this possibility.

You might also want to check whether the effects that appear active could be highly correlated with other effects. When this occurs, one effect can mask the true significance of another effect. The Color Map in Figure 8 displays absolute correlations between effects.

2. Click **Make Model**.

   The constructed model contains only the effects X1, X11, and X12.

3. Click **Run** in the Model Specification window.

Figure 7 Parameter Estimates for Model

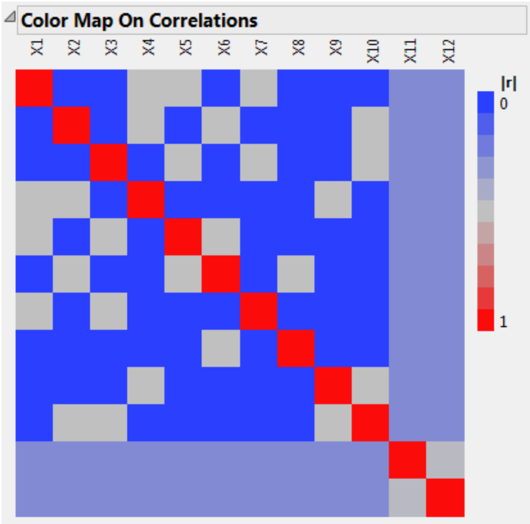
Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	100.12645	0.319502	313.38	<.0001*
X11	11.347787	0.360447	31.48	<.0001*
X1	9.8209584	0.319502	30.74	<.0001*
X12	-1.1329	0.360447	-3.14	0.0347*

Note that the parameter estimates for X11 and X1 are close to the theoretical values that you used to simulate the model. See Figure 4, where you specified a model with X1 = 10 and X11 = 10. The significance of the factor X12 is an example of a false positive.

4. In your Custom Design window, open the **Design Evaluation > Color Map on Correlations** outline.

Figure 8 Color Map on Correlations Outline



With your cursor, place your mouse pointer over cells to see the absolute correlations. Notice that X1 has correlations as high as 0.5 with other main effects (X4, X5, X7). (Figure 8 uses JMP default colors.)

Analyze a Supersaturated Design Using Stepwise Regression

Stepwise regression is another way to identify active factors. The design table in Figure 5 contains three scripts. The Model script analyzes your data using stepwise regression in the Fit Model platform.

1. In the Tables panel of the design table, select **Run Script** from the red triangle next to **Model**.
2. Change the **Personality** from **Standard Least Squares** to **Stepwise**.
3. Click **Run**.
4. In the Stepwise Fit for Y report, change the **Stopping Rule** to **Minimum AICc**.

   For designed experiments, BIC is typically a more lenient stopping rule than AICc as it tends to allow inactive effects into the model.
5. Click **Go**.

Figure 9 Stepwise Regression for Supersaturated Design

Stepwise Fit for Y

Stepwise Regression Control

Stopping Rule: Minimum AICc
Direction: Forward

Enter All | Make Model | Remove All | Run Model | Go | Stop | Step

SSE	DFE	RMSE	RSquare	RSquare Adj	Cp	p	AICc	BIC
10.303713	5	1.4355287	0.9950	0.9930	.	3	46.06085	33.04528

Current Estimates

Lock	Entered	Parameter	Estimate	nDF	SS	"F Ratio"	"Prob>F"
☑	☑	Intercept	100.288292	1	0	0.000	1
☐	☑	X1	9.65911557	1	696.6289	338.048	8.75e-6
☐	☐	X2	0	1	6.663958	7.324	0.05375
☐	☐	X3	0	1	0.025784	0.010	0.92503
☐	☐	X4	0	1	1.069475	0.463	0.53346
☐	☐	X5	0	1	0.367431	0.148	0.72011
☐	☐	X6	0	1	1.775829	0.833	0.41305
☐	☐	X7	0	1	1.01185	0.436	0.54533
☐	☐	X8	0	1	1.087016	0.472	0.52993
☐	☐	X9	0	1	1.130717	0.493	0.52127
☐	☐	X10	0	1	1.039842	0.449	0.53951
☐	☑	X11	10.8622582	1	825.9206	400.788	5.75e-6
☐	☐	X12	0	1	7.334066	9.879	0.03475

Step History

Step	Parameter	Action	"Sig Prob"	Seq SS	RSquare	Cp	p	AICc	BIC	
1	X11	Entered	0.0147	1354.353	0.6570	.	2	70.555	64.7933	○
2	X1	Entered	0.0000	696.6289	0.9950	.	3	46.0609	33.0453	○
3	X12	Entered	0.0347	7.334066	0.9986	.	4	54.775	25.1722	○
4	X2	Entered	0.0374	2.405438	0.9997	.	5	97.4888	13.9655	○
5	X10	Entered	0.0362	0.524072	1.0000	.	6	.	-5.1002	○
6	X5	Entered	0.2140	0.035771	1.0000	.	7	.	-20.77	○
7	Best	Specific	.	.	0.9950	.	3	46.0609	33.0453	●

Figure 9 shows that the selected model consists of the two active factors, X1 and X11. The step history appears in the bottom part of the report. Keep in mind that correlations between X1 and X11 and other factors could mask the effects of other active factors. See Figure 8.

Note: This example defines two large main effects and sets the rest to zero. Real-world situations can be less likely to have such clearly differentiated effects.