

Classical Designs: Full Factorial Designs

Authored by: Cory Natoli

Steve Oimoen, PhD

1 June 2019



The goal of the STAT COE is to assist in developing rigorous, defensible test strategies to more effectively quantify and characterize system performance and provide information that reduces risk. This and other COE products are available at www.afit.edu/STAT.

Table of Contents

Executive Summary.....	2
Introduction	2
Factorial Basics.....	3
Properties.....	5
2 ⁴ Factorial Design Matrix	7
Testing for Curvature	8
Center Points.....	8
Checking for Curvature	9
DOE Fundamentals	10
Randomization	10
Replication	11
Blocking.....	12
Conclusion.....	14
References	16
JMP Demo	17

Executive Summary

Classical designs are a common starting point for the design phase of testing. A full factorial design is a simple systematic design style that allows for estimation of main effects and interactions. This design is very useful, but requires a large number of test points as the levels of a factor or the number of factors increase. Assessing the tradeoff between budget and the information gained in a full factorial design is an important consideration to take into account. However, the full factorial design does have a number of properties that make it very powerful. This paper discusses how to design an effective full factorial design and describes when it is appropriate to use. Additionally, a demo using the statistical software package JMP provides an example.

Keywords: Randomization, blocking, main effects, interactions, experimental design, JMP

Preface

Design is a key step in the Scientific Test and Analysis Technique (STAT) process; however, it is important that the steps leading to this phase are followed properly to ensure meaningful results. Before designing a test, the system must be appropriately decomposed and a list of factors and responses should be generated. Additionally, constraints (of which some will be discussed in this paper) must be identified in order to choose the best design option. For more information on the STAT process, see our “Guide to Developing an Effective STAT Test Strategy” on our website www.afit.edu/STAT (Burke et al. 2018).

Introduction

Testing incorporates four distinct phases: Plan, Design, Execute, and Analyze (see Figure 1).

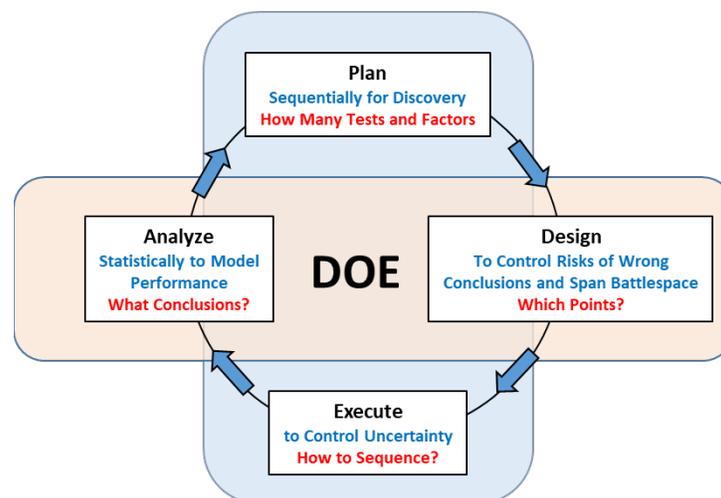


Figure 1: Four Distinct Testing Phases

The Plan phase is crucial for identifying the test objectives, responses and factors, as well as their ranges and levels, and the size of a test design that is feasible for execution. A good design relies on an informed planning phase and seeks to balance the tradeoffs among budget, risk, and information gained. A classical design is a common starting point test design construction. Classical designs include full factorial and fractional factorial designs. A special case of the full factorial design is the 2^k factorial design, which has k factors where each factor has just two levels. A full factorial design consists of all possible factor combinations in a test, and, most importantly, varies the factors simultaneously rather than one factor at a time. Using this approach, the tester can examine both main effects (effect of the independent variables on the dependent variable) and interactions (effect of the interaction between independent variables on the dependent variable) associated with both categorical and continuous factors. The number of effects in a 2^k factorial design:

- k main effects
- $\binom{k}{2} = \frac{k(k-1)}{2}$ 2-factor interactions
- $\binom{k}{3} = \frac{k(k-1)(k-2)}{6}$ 3-factor interactions
- $\binom{k}{n} = \frac{k!}{n!(k-n)!}$ n -factor interactions, $n \leq k$
- 1 k -factor interaction

However, full factorial designs do require a larger sample size as the number of factors and associated levels increase. For these reasons, full factorial designs may allow you to estimate every possible interaction, although you are probably only interested in two-factor interactions or possibly three-factor interactions. Lastly, this design may not be the best approach when there is a strict limit on the number of test runs due to constraints such as time, cost, resources, the number of factors and levels, and randomization. However, it can be useful to utilize general factorial designs in limited circumstances such as dealing with multi-level categorical factors or when the experiment only has two to three factors. While other full factorial designs such as a general factorial design do exist, this paper discusses the 2^k factorial design.

Factorial Basics

A 2^k factorial design follows a pattern that captures every possible combination of factor and level in the design matrix. This approach builds a model that estimates all of the main effects and each interaction. Table 1 shows an example 2^k factorial design of three factors, each with two levels, coded as low (-) and high (+).

Table 1: 2³ Full Factorial Design Matrix

Run	Factor		
	A	B	C
1	-	-	-
2	+	-	-
3	-	+	-
4	+	+	-
5	-	-	+
6	+	-	+
7	-	+	+
8	+	+	+

Replacing the coding with actual values for any specific test is straightforward. For instance, if one of the factors in the design was distance that ranged from 500 to 1000 feet, then the “-” level would be 500 feet and the “+” level would be 1000 feet. The order created using the systematic pattern shown above is *standard order*. This ordering scheme introduces one factor at a time and then combines with the factors preceding it. In practice, randomize test execution to ensure independence and guard against hidden factors. Failure to randomize may identify significant factor effects that are actually due to uncontrolled variables. The visual representation of the design space for a 2³ factorial design is a geometric cube where the test points are “mapped” to each corner of the cube. Figure 2 provides a graphical representation of this cube with coding “-1” for the low setting and “1” for the high setting of each factor.

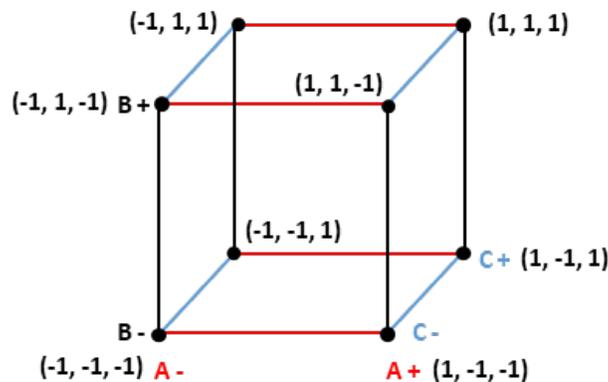


Figure 2: 2³ Full Factorial Design

The properties of the 2^k factorial design matrix are:

- Every column has an equal number of - and + signs (low and high settings)
- The sum of the product of signs in any two columns is zero (called the orthogonality property)
- Multiplying any column by 1 leaves that column unchanged (identity property)
- Product of any two columns yields a column that can calculate the effect of another term (for example, $A \times B = AB$)

Properties

The three primary properties of all factorial designs are estimable model terms, projection, and orthogonality. Using a factorial design, the experiment examines all possible combinations of levels for each factor. Since every combination of factor and level is included in the 2^k factorial design, the 2^3 design consists of unique columns for every main effect, every two-factor interaction effect and the three-factor interaction effect. Larger values of k will include many more interactions from two-factor all the way up to a k -factor interaction. Table 2 lists all main effects, two-factor interactions, and the three-factor interaction effects for the 2^3 factorial design.

Table 2: 2^3 Factorial Design with Interactions

Run	Effects						
	A	B	C	AB	AC	BC	ABC
1	-1	-1	-1	1	1	1	-1
2	1	-1	-1	-1	-1	1	1
3	-1	1	-1	-1	1	-1	1
4	1	1	-1	1	-1	-1	-1
5	-1	-1	1	1	-1	-1	1
6	1	-1	1	-1	1	-1	-1
7	-1	1	1	-1	-1	1	-1
8	1	1	1	1	1	1	1

The projection property states that factorial designs “project” into replicated factorial designs in lower dimensions. The STAT COE Best Practice, “Classical Designs: Fractional Factorial Designs” discusses this particular topic in more detail (Natoli, 2018). Figure 3 shows a 2^3 factorial design (factors A, B, and C).

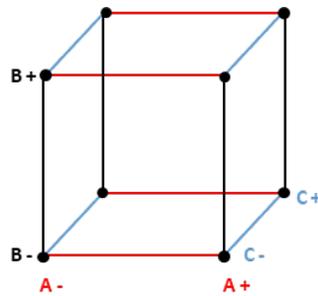


Figure 3: 2^3 Factorial Design

If one of those factors is insignificant, the design can collapse on the insignificant factor (see Figure 4). This process then allows for replicates of the remaining factors. This property is particularly useful for sequential testing, where follow-on testing has been planned.

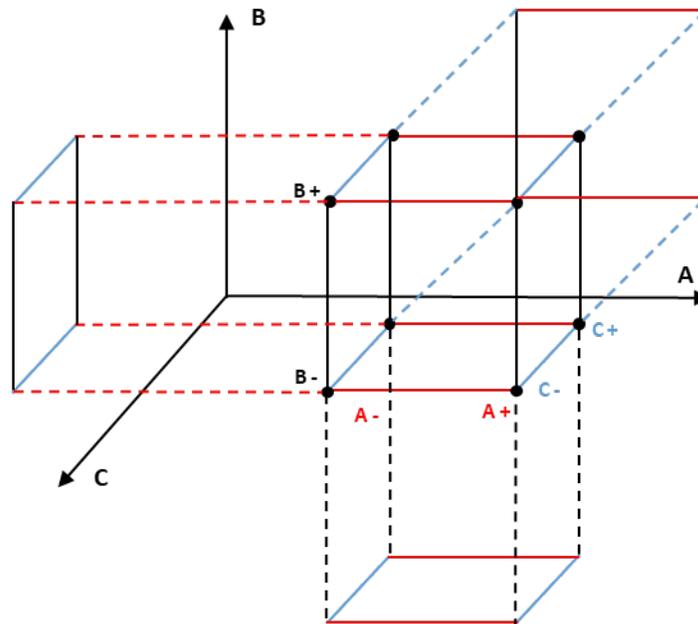


Figure 4: Three-factor Projection

All 2^k factorial designs are also orthogonal designs. Orthogonality means the model terms are uncorrelated with one another and therefore estimated independently of each other. This feature helps clearly determine which factor affects the response. Without orthogonality, the process to determine which factor is truly contributing to the response is more complex. Having the design matrix contain all possible combinations of factors and levels in your test provides orthogonality and:

- A balanced, symmetric design
- Independent estimates of main effects and ALL interaction effects
 - Zero correlation between estimates of the effects

- When removing one effect from the model, the estimates of remaining effects are unchanged
- The ability to project into more effective designs when insignificant effects exist

2⁴ Factorial Design Matrix

Table 3 shows an example of a 2⁴ factorial design. The table consists of plus and minus signs and includes columns for every main effect, two-factor, three-factor, and a four-factor interaction effect. All model terms (the main effects of each factor and all interactions) could be estimated with this design. Typically, however, main effects and two-factor interactions are sufficient to describe the variability in the measured response. The table highlights the factor settings of the executable design in blue.

Table 3: Design and Model Matrix – 4 Factors

Run	I	A	B	C	D	AB	AC	AD	BC	BD	CD	ABC	ABD	BCD	ACD	ABCD
1	1	-	-	-	-	+	+	+	+	+	+	-	-	-	-	+
2	1	+	-	-	-	-	-	-	+	+	+	+	+	-	+	-
3	1	-	+	-	-	-	+	+	-	-	+	+	+	+	-	-
4	1	+	+	-	-	+	-	-	-	-	+	-	-	+	+	+
5	1	-	-	+	-	+	-	+	-	+	-	+	-	+	+	-
6	1	+	-	+	-	-	+	-	-	+	-	-	+	+	-	+
7	1	-	+	+	-	-	-	+	+	-	-	-	+	-	+	+
8	1	+	+	+	-	+	+	-	+	-	-	+	-	-	-	-
9	1	-	-	-	+	+	+	-	+	-	-	-	+	+	+	-
10	1	+	-	-	+	-	-	+	+	-	-	+	-	+	-	+
11	1	-	+	-	+	-	+	-	-	+	-	+	-	-	+	+
12	1	+	+	-	+	+	-	+	-	+	-	-	+	-	-	-
13	1	-	-	+	+	+	-	-	-	-	+	+	+	-	-	+
14	1	+	-	+	+	-	+	+	-	-	+	-	-	-	+	-
15	1	-	+	+	+	-	-	-	+	+	+	-	-	+	-	-
16	1	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+

Design

Model = Linear Effects + Interactions

A factor must have at least two levels in order to determine whether changing that factor causes a change in the response. However, additional factors and factor levels rapidly increase the size of the full factorial design. This is often the downside to this design approach. When a test has more than four factors, a full factorial design may contain more runs than necessary.

Testing for Curvature

Center Points

Unfortunately, 2-level designs assume the effect of the factors on the response is linear, which may not be a good assumption. To estimate a line, two points are necessary; whereas to estimate a curve, three points are required (see Figure 5).

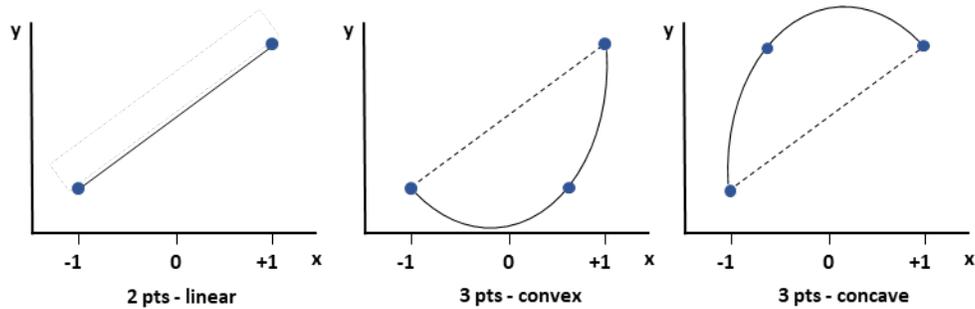


Figure 5: Points Required Estimating a Line and Curve

However, the tester can add center points to the original design to test this assumption. Center points are additional test points where each factor is set to the midpoint of its range. In coded units, this means that every continuous factor is set to 0. For example, Figure 6 shows all of the possible combinations of a 2^4 factorial design where all factors are continuous. The red points in the figure represent the center points, which can be used to determine whether the assumption of linear effects on the response is reasonable.

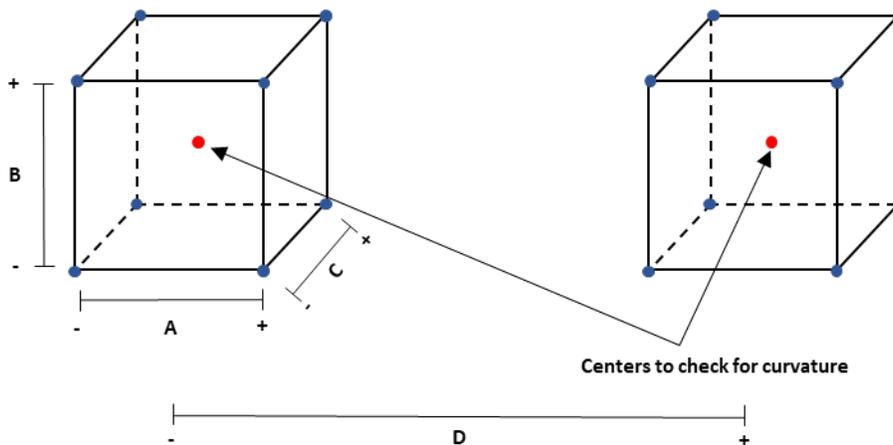


Figure 6: 2^4 Full-factorial Design with Center Points

When the test contains categorical factors, you must use pseudo-center points since there is no “midpoint” for a categorical factor. For example, in a 2^3 design with two continuous factors and one categorical factor (see Figure 7), we can add two center points at one level of the categorical factor and

two center points at the other level of the categorical factor. Since Location is a categorical factor, there is no halfway point along a categorical axis, so pseudo-center points are required. When adding pseudo-center points, keep the number of continuous center points balanced for each level of the categorical factor.

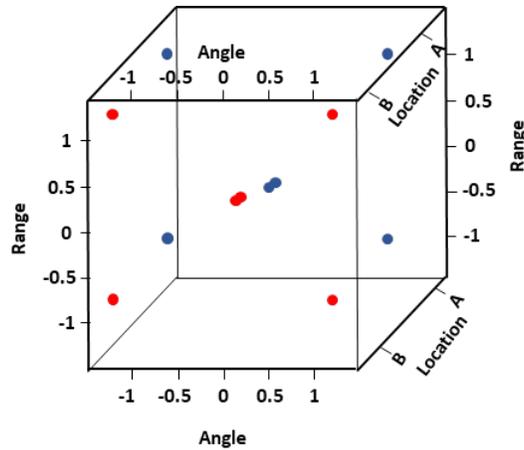


Figure 7: Center Points and Qualitative Factors

Center points are useful to include in the initial design or as an augment (addition after initial testing) to the design if a linear model does not fit well. Typically, place center points non-randomly. Put a center point at the beginning, middle and the end of the experiment to assist in checking for some systematic bias or drift in the response value. In addition to checking for curvature (explained in the next section), there are several additional practical uses of including center points in the design, including:

- Use current operating conditions as the center point for a baseline response
- Check for “abnormal” conditions during the time the experiment is conducted
 - This checks to make sure system performance is as expected on a given day
- Check for time trends
 - Spacing the center points at the beginning, middle, and end can show a difference due to time
- Check for non-linearity in the response (discussed in the next section)
- Use in first few runs (there is little or no information available about the magnitude of error)
 - This provides an estimate for the “noise” in the system that can be used for sizing a design

Checking for Curvature

Adding center points to a two-level design provides an estimate of pure error and the ability to test for lack-of-fit. Pure error is due to variability in the measurement of the dependent variable and is one component of the residual error. Subtracting the pure error from the residual error estimates in lack-of-fit of the model. The lack-of-fit test compares the average response of the center points with the

average response of the factorial points (see Figure 8). A significant difference between the two point values indicates the linear model is inadequate, clearly demonstrating significant lack-of-fit. When there is significant lack-of-fit, the model is not sufficiently capturing the variability in the response. There is behavior in the response that is not being explained by the current model.

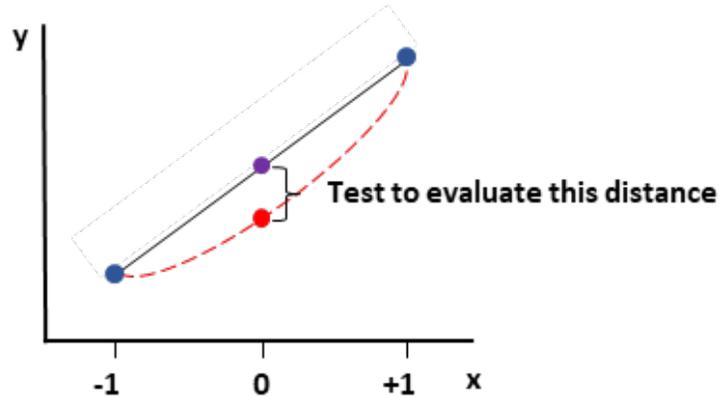


Figure 8: Checking for Curvature

Adding center points to a two-level design does not allow estimation of specific quadratic effects; we cannot determine which factor has a quadratic effect on the model. However, center points do provide an indication if any quadratic terms are needed in the model. To adjust the model, additional test points would need to be executed in order to determine which factor(s) have a quadratic effect on the response. We recommend including three to five center points in the design to test for curvature.

DOE Fundamentals

DOE has three foundational principles: randomization, replication, and blocking. We discuss how to implement these three principles when executing a 2^k design in the following subsections.

Randomization

The first principle, randomization, executing test runs in a random order, supports the use of statistical methods. Randomly ordering the actual experimental runs helps to validate the assumption of independently distributed observations (Montgomery, 2017). Randomization is one of the easiest ways to control for noise in an experiment. Randomizing protects against known and unknown events happening in the background, neither of which are controllable. Executing runs in random order helps “average out” the effects of “lurking” variables, which improves the estimation of the effects of the other factors. Figure 9 shows all possible combinations of three factors at two levels each. The first approach shows standard order with a structured sequence. This approach can have outside influences over time, while using a randomized order averages out outside influences. For instance, because the “C” factor is not changed randomly, being able to determine whether a change in the response was due to changing the “C” factor levels or due to some outside lurking variable is not possible.

Run	A	B	C	Response
1	-1	-1	-1	
2	1	-1	-1	
3	-1	1	-1	
4	1	1	-1	
5	-1	-1	1	
6	1	-1	1	
7	-1	1	1	
8	1	1	1	
9	0	0	0	
10	0	0	0	

Standard Order

Run	A	B	C	Response
2	1	-1	-1	
8	1	1	1	
5	-1	-1	1	
10	0	0	0	
4	1	1	-1	
1	-1	-1	-1	
3	-1	1	-1	
7	-1	1	1	
9	0	0	0	
6	1	-1	1	

Randomized

Figure 9: Standard and Random Run Order

Important to note is *if you cannot fully randomize the runs because it is too expensive or time intensive to do so, a 2^k design is not the best test strategy*. Other statistical design approaches are available to use in these circumstances (Montgomery, 2017).

Replication

The second foundational principle of DOE, replication, is repeating each factor/level combination independently of any previous run. Two significant attributes of replication are being able to determine an estimate of the experimental error (used to decide if noted differences in the data are actually statistically different) and a more precise estimate of the actual parameter value (Montgomery, 2017). Without replication, determining why two data points differ is not possible. As previously discussed, a 2^k factorial design can support a model with $2^k - 1 = n - 1$ terms (the intercept term, main effects, and all interaction terms). Model terms are estimated at the cost of a run, which means that you will need more runs than model terms in order to estimate the effects and determine whether the effect is statistically significant or not. The number of runs in the design must be greater than the number of parameters in the model in order to perform hypothesis testing on each model parameter. Consider

using replication in order to perform hypothesis tests, obtain an estimate of “pure error”, and increase the power of the design.

Blocking

Blocking is the third DOE principle used to minimize or remove potential variability due to a *known* nuisance factor. This planning and analysis technique sets aside (blocks out) unwanted and known nuisance variability. The benefit of incorporating a blocking approach is reduced experimental error, which in turn increases the power to identify the impact of test conditions. If the nuisance factor is a source of noise, blocking will reduce the influence. By not blocking on a nuisance factor, detecting whether a factor is significant becomes more difficult. Replacing replicates with blocks and then randomizing within each block is a Randomized Completed Block Design (RCBD). Each block contains a complete replicate of the factor setting. Randomizing in this manner maintains the orthogonality of the factors (see Figure 10).

Run	X1	X2	X3	Block
1	1	1	1	1
2	-1	1	1	1
3	1	-1	1	1
4	-1	-1	1	1
5	1	1	-1	1
6	-1	1	-1	1
7	1	-1	-1	1
8	-1	-1	-1	1
9	1	1	1	2
10	-1	1	1	2
11	1	-1	1	2
12	-1	-1	1	2
13	1	1	-1	2
14	-1	1	-1	2
15	1	-1	-1	2
16	-1	-1	-1	2

Figure 10: Blocking a replicated 2^k design

When you cannot afford replicates, you must sacrifice the ability to estimate a high order interaction in order to account for the blocking effect. Consider a 2^4 factorial design. Construct two groups of eight, choosing blocks confounded (meaning the tester will not be able to differentiate that effect from the blocking effect) with the highest-order interaction “ABCD” (see

Table 4). Place all “+1s” of the highest order interaction in one block and all “-1s” of the same interaction in the other block. After constructing both blocks, randomize the run order within each block.

Table 4: Four-factor Full Factorial Design

Run	A	B	C	D	AB	AC	AD	BC	BD	CD	ABC	ABD	BCD	ACD	ABCD
1	-	-	-	-	+	+	+	+	+	+	-	-	-	-	+
2	+	-	-	-	-	-	-	+	+	+	+	+	-	+	-
3	-	+	-	-	-	+	+	-	-	+	+	+	+	-	-
4	+	+	-	-	+	-	-	-	-	+	-	-	+	+	+
5	-	-	+	-	+	-	+	-	+	-	+	-	+	+	-
6	+	-	+	-	-	+	-	-	+	-	-	+	+	-	+
7	-	+	+	-	-	-	+	+	-	-	-	+	-	+	+
8	+	+	+	-	+	+	-	+	-	-	+	-	-	-	-
9	-	-	-	+	+	+	-	+	-	-	-	+	+	+	-
10	+	-	-	+	-	-	+	+	-	-	+	-	+	-	+
11	-	+	-	+	-	+	-	-	+	-	+	-	-	+	+
12	+	+	-	+	+	-	+	-	+	-	-	+	-	-	-
13	-	-	+	+	+	-	-	-	-	+	+	+	-	-	+
14	+	-	+	+	-	+	+	-	-	+	-	-	-	+	-
15	-	+	+	+	-	-	-	+	+	+	-	-	+	-	-
16	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+

Then randomize run order within blocks!

When blocking an un-replicated design, at least one model term will be confounded with the block. The sparsity of effects principle states a subset of main effects and low-order interactions dominate most systems. In other words, main effects and two-factor interactions typically account for most of the variability in the response. Higher order interactions (like 3-factor or 4-factor interactions) are rarely needed to explain the response. When blocking, ensure the highest-order interaction effect or an interaction believed not to be significant is confounded. This approach ensures systematically designed tests for all blocks to facilitate estimates of effects within each block.

Conclusion

This paper discussed the power of full factorial designed experiments, their tradeoffs, and 2^k designs. Randomization, replication, and blocking are the three foundational DOE principles. A full factorial design test is useful for estimating main effects and interactions by varying the factors together, but can grow large with an increasing numbers of factors. These designs have a number of properties that allow for effective model building. A full factorial design is a good starting point for design discussion and can serve as a baseline design in a number of different test scenarios. Use general factorial designs in rare cases such as having categorical factors or only having two to three factors. Avoid using the full factorial

design if the number of runs is restricted, design region constraints exist, or the number of factors is more than four.

References

Burke, Sarah et al. "Guide to Developing an Effective STAT Test Strategy V6.0." Scientific Test and Analysis Techniques Center of Excellence (STAT COE), 31 Dec. 2018.

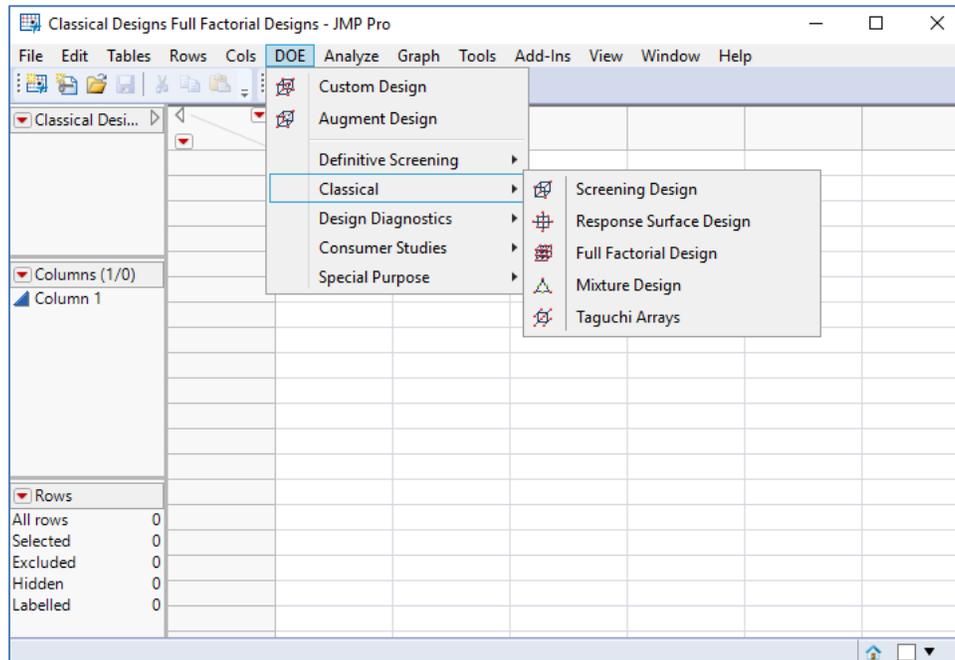
Montgomery, Douglas C. *Design and Analysis of Experiments*. 9th ed., John Wiley & Sons, Inc., 2017.

Natoli, Cory. "Classical Designs: Fractional Designs." Scientific Test and Analysis Techniques Center of Excellence (STAT COE), 30 Sep. 2018.

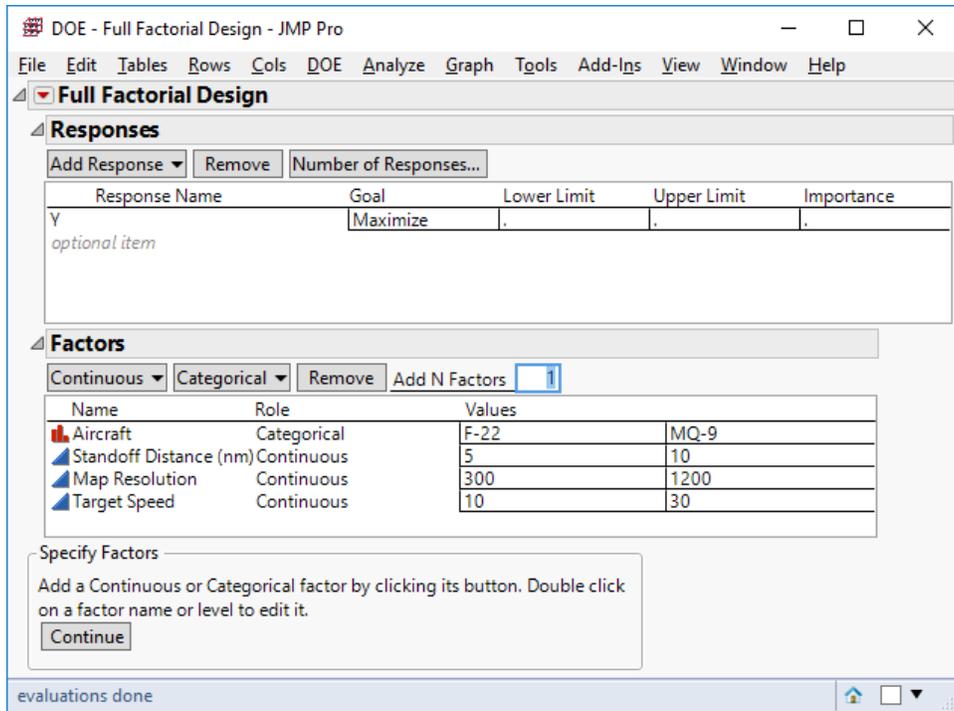
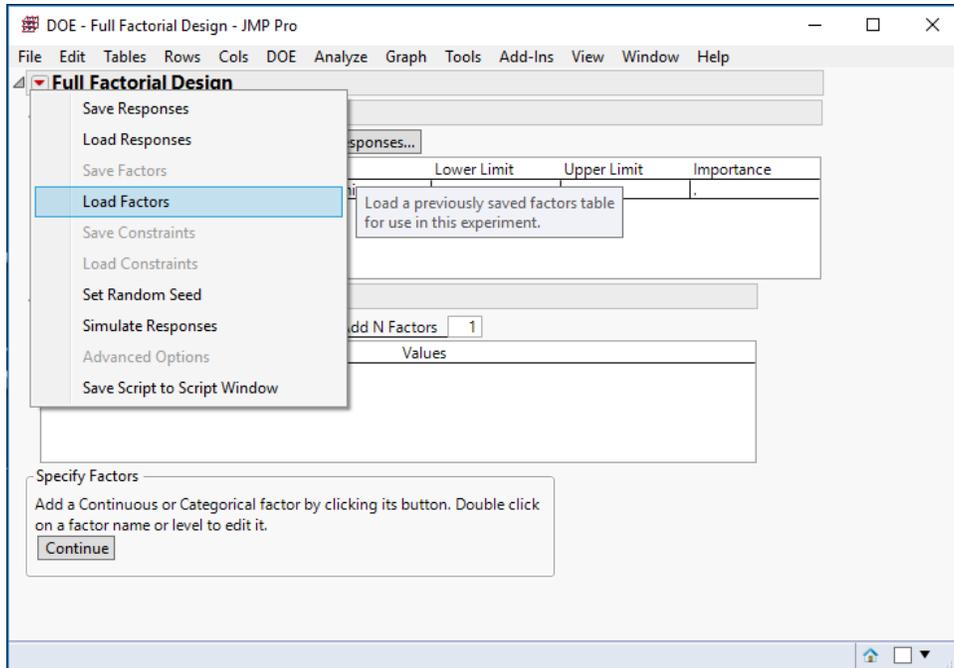
JMP Demo

This example will create a 2^4 factorial design for a test with one categorical and three continuous factors using JMP V. 13. The categorical factor is Aircraft and has two levels (F-22 and MQ-9). The continuous factors are Standoff Distance (nautical miles, 5, 10), Map Resolution (dots per inch (dpi), 300, 1200), and Target Speed (knots, 10, 30). The response is Time to Locate (seconds) and the goal is to “Minimize” the response Time to Locate.

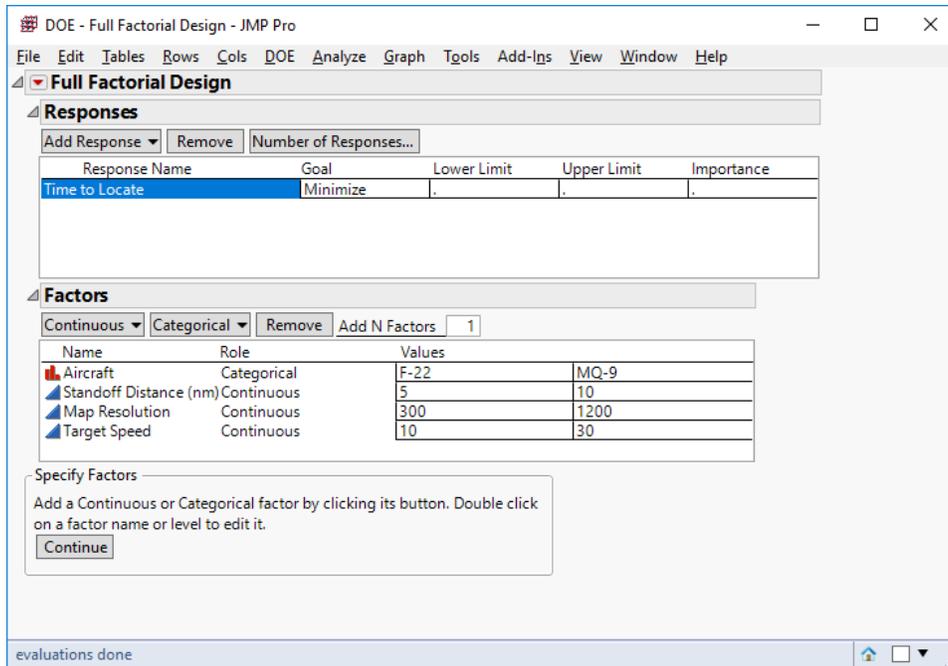
- 1) Select “DOE -> Classical -> Full Factorial Design”.



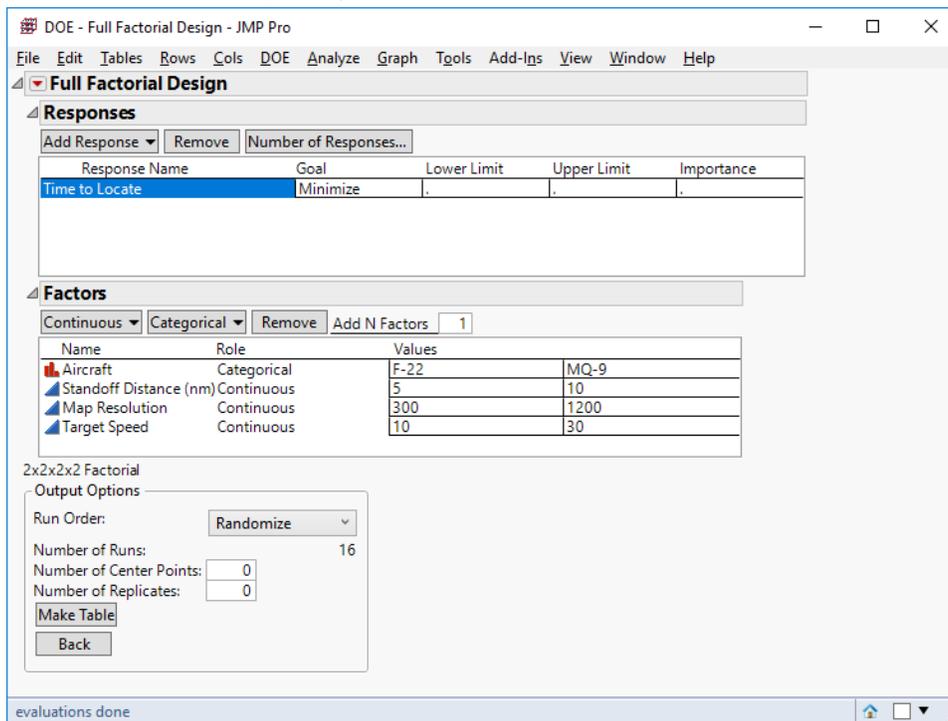
- 2) Enter the factors into the factor section.



- Update the Response Name to "Time to Locate" and the goal to "Minimize", and then select "Continue".



- 4) The number of runs will be $2^4 = 16$ with no center points (although you could add replicates or center at this menu if desired). Be sure to select Run Order as “Randomize” and then click “Make Table”. Remember, always randomize!



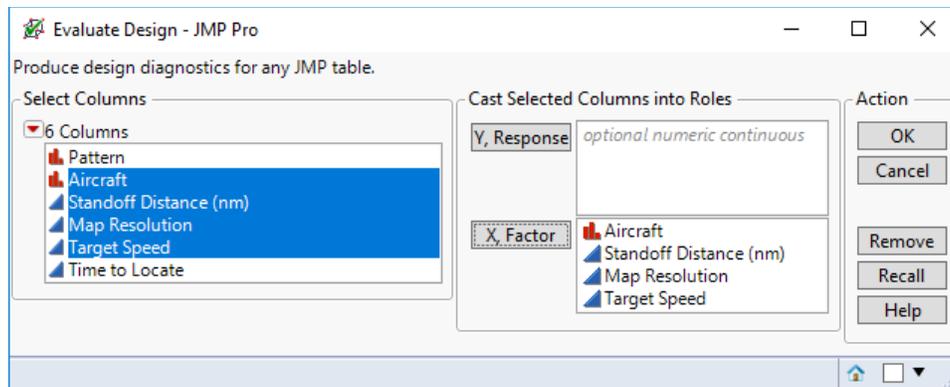
- 5) The design table shows the settings for each of the factors in each run. A column for the response is included which you can fill in as you execute the test.

Run	Pattern	Aircraft	Standoff Distance (nm)	Map Resolution	Target Speed	Time to Locate
1	2+++	MQ-9	10	300	30	
2	2---	MQ-9	5	300	30	
3	1---	F-22	5	300	30	
4	1+++	F-22	10	300	10	
5	1---	F-22	10	300	30	
6	2---	MQ-9	10	300	10	
7	2---	MQ-9	5	1200	10	
8	2+++	MQ-9	10	1200	30	
9	1+++	F-22	10	1200	30	
10	2+++	MQ-9	10	1200	10	
11	1+++	F-22	10	1200	10	
12	1---	F-22	5	300	10	
13	1---	F-22	5	1200	10	
14	2---	MQ-9	5	300	10	
15	2+++	MQ-9	5	1200	30	
16	1+++	F-22	5	1200	30	

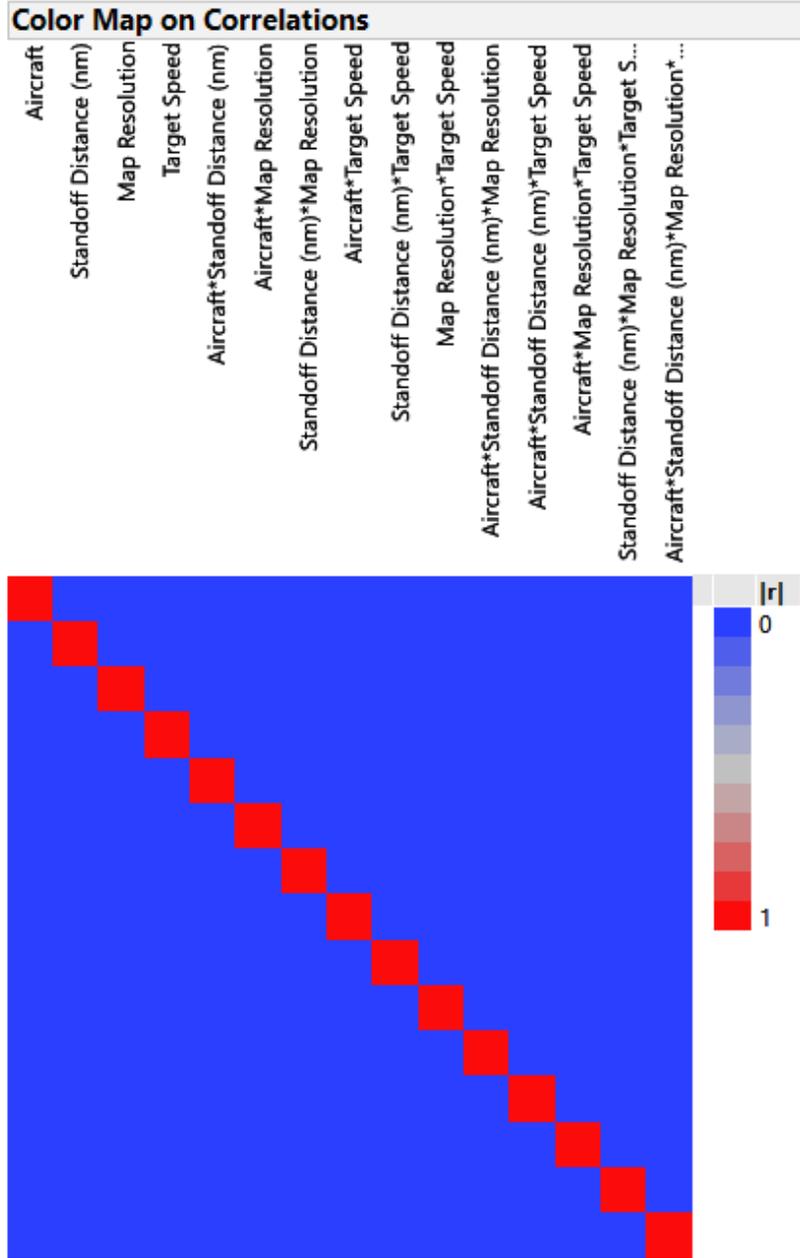
- 6) To view correlation between any two effects, select “DOE -> Design Diagnostics -> Evaluate Design”.

Run	Pattern	Aircraft	Standoff Distance (nm)	Map Resolution	Target Speed	Time to Locate
7	2+-	MQ-9	5	1200	10	
8	2+-	MQ-9	10	300	10	
9	2---	MQ-9	5	300	10	
10	2---	MQ-9	5	300	30	
11	1+++	F-22	5	1200	30	
12	1+++	F-22	10	1200	30	
13	1+++	F-22	10	1200	10	
14	1+-	F-22	5	1200	10	
15	1+++	F-22	10	300	30	
16	2---	MQ-9	5	1200	30	

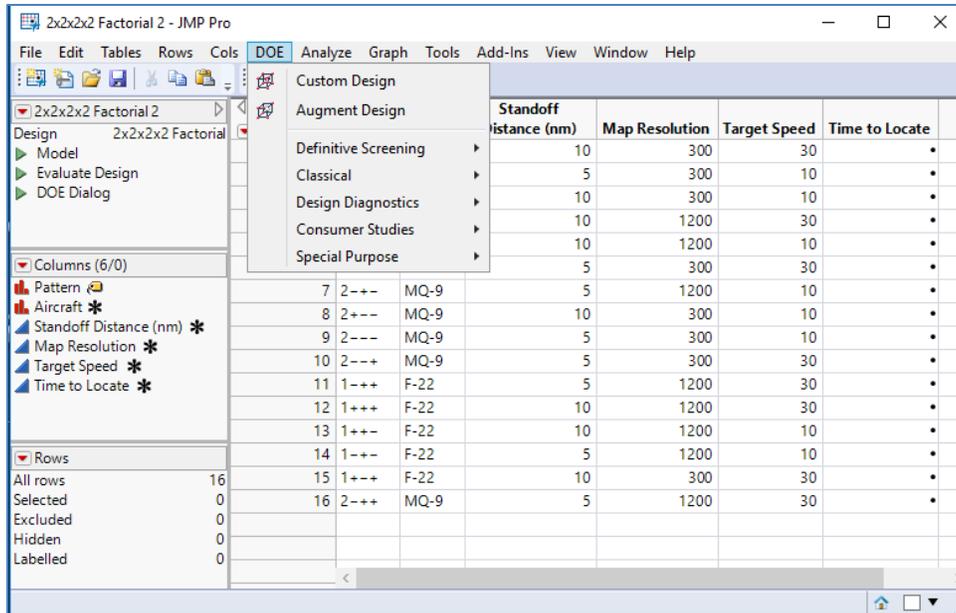
- 7) Load the factors into the “X, Factor” box and select “OK”.



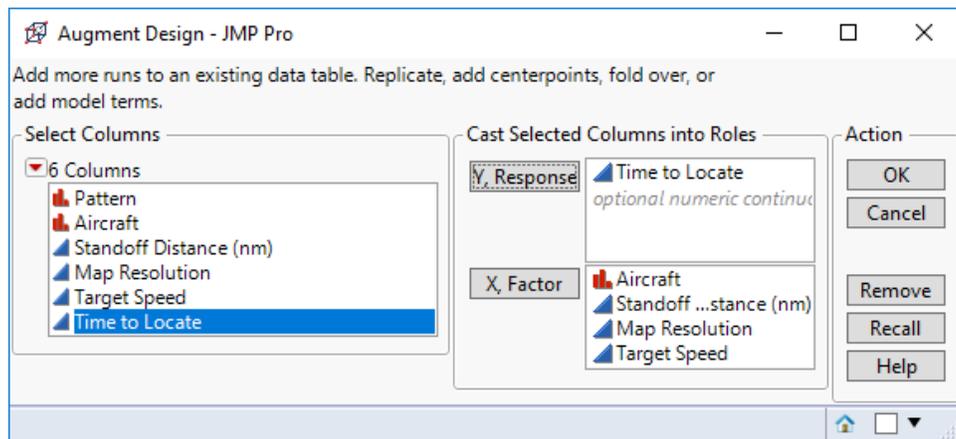
- 8) Scroll down and select the drop down next to “Color Map on Correlations”. Note that because there is blue all in the off-diagonal, there is no aliasing of effects in this design. This means the design is orthogonal (as expected) and we will be able to estimate each model term. Also, note there are four main effects, six two-factor interactions, four three-factor interactions, and one four-factor interaction.



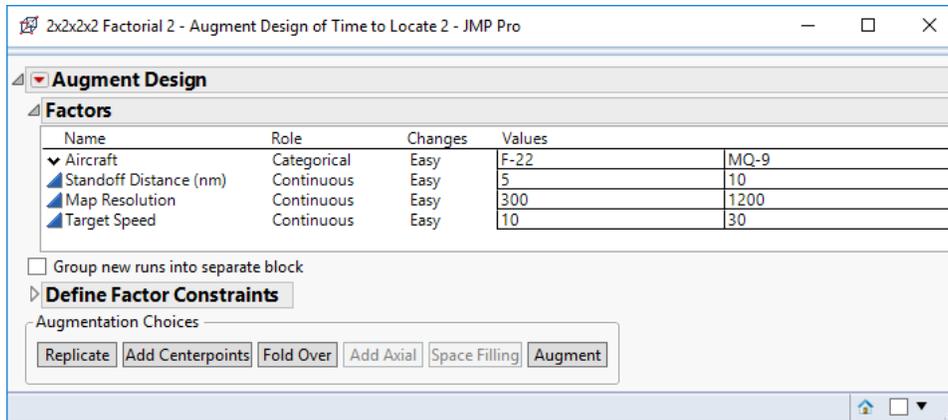
- 9) If additional runs are required (possibly some center points), from the design data table, select “DOE -> Augment Design”.



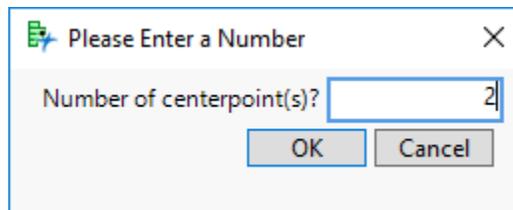
10) Load all of the factors into the “X, Factor” box and the response “Time to Locate” into the “Y, Response” box. Select “OK”.



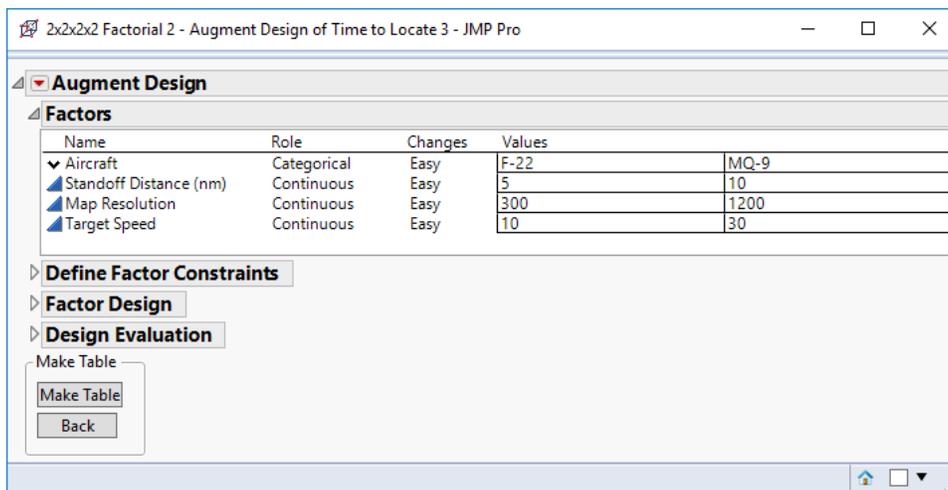
11) Select “Add Centerpoints”.



12) Enter the number of centerpoints (for this example, 2 are added), then select “OK”.



13) Next, select “Make Table”



14) The new augmented design table now has 18 runs. The first 16 runs represent the original factorial design with the additional center points added at the end.

The screenshot shows the JMP Pro interface for an augmented design. The main data table is as follows:

	Aircraft	Standoff Distance (nm)	Map Resolution	Target Speed	Time to Locate
1	MQ-9	10	300	30	•
2	F-22	5	300	10	•
3	F-22	10	300	10	•
4	MQ-9	10	1200	30	•
5	MQ-9	10	1200	10	•
6	F-22	5	300	30	•
7	MQ-9	5	1200	10	•
8	MQ-9	10	300	10	•
9	MQ-9	5	300	10	•
10	MQ-9	5	300	30	•
11	F-22	5	1200	30	•
12	F-22	10	1200	30	•
13	F-22	10	1200	10	•
14	F-22	5	1200	10	•
15	F-22	10	300	30	•
16	MQ-9	5	1200	30	•
17	F-22	7.5	750	20	•
18	MQ-9	7.5	750	20	•

Summary statistics from the interface:

- All rows: 18
- Selected: 0
- Excluded: 0
- Hidden: 0
- Labelled: 0